Projecting contact matrices in 177 geographical regions: an update and comparison with empirical data for the COVID-19 era

Supplementary Material

Kiesha Prem, Kevin van Zandvoort, Petra Klepac, Rosalind M Eggo, Nicholas G Davies,
Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group,
Alex R Cook, Mark Jit*

Contents

1 Materials and methods

1.1 Population age structure

The United Nations Population Division provides the population age composition in five-year age intervals P_a (i.e.,number of individuals in age group a) for the 177 geographical regions in this study [1]. They also present the estimation of rural and urban population age compositions for the 177 geographical regions [2].

---

*Correspondence to mark.jit@lshtm.ac.uk.

1.2 Household age structure

1.2.1 Household data

We extracted the household data from POLYMOD contact study [3]—for the eight POLYMOD countries: Belgium, Germany, Finland, the United Kingdom, Italy, Luxembourg, the Netherlands, and Poland. The Demographic Household Surveys (DHS) provides nationally-representative household surveys of the whole population for several low- and middle-income countries [4]. Table 1 lists the 43 DHS countries included in the study. The largest survey conducted in India in 2015 included ~3 million individuals from ~600,000 households. The DHS also provides data for rural and urban areas, allowing us to derive rural-urban household age matrices.

1.2.2 Household age matrix (HAM)

To estimate the age-specific contacts at home for county c, we need to first construct the household age matrix (HAM) for country c. The HAM of country c, $\mathbf{h}_{a,\alpha}^c$, represents the mean number of household members of age α of an individual aged a. Although this can be easily derived for the 8 POLYMOD and 43 DHS countries, it is not straightforward for other countries or geographical regions without available household structure data. In Prem et al. [5], we provide detailed steps to project a country's household age structure.

1. **Population ratio matrices:** For the 51 POLYMOD and DHS countries, we first derive their population ratio matrices $\mathbf{\Pi}_{a,\alpha}^c$ by dividing the elements of HAM ($\mathbf{h}_{a,\alpha}^c$) by the proportion of the population of country c aged a, P_a^c. The population ratio matrix of a country measures the propensity of having a household member aged α for an individual aged a after adjusting for the population age structure. Because we know the population age composition for the 177 geographical regions, we can project their HAM by extrapolating the relationship between P_a^c and $\mathbf{h}_{a,\alpha}^c$ as estimated from the POLYMOD and DHS countries, and applying that to the country's population age profile. We could simply use the mean population ratio matrix of the POLYMOD and DHS countries to derive the HAM of country c, $\mathbf{h}_{a,\alpha}^c$, as follows $P_a^c \overline{\mathbf{\Pi}_{a,\alpha}^c}$.

2. **HAM weights:** However, as household age structures vary across countries in different stages of development and with different demographics, we use 14 country characteristics from the World Bank and United Nations Educational, Scientific and Cultural Organization Institute for Statistics (UIS) databases [6] to quantify the similarity of these 126 geographical regions (with no available household data). The 14 country characteristics include gross domestic product per capita, total fertility rate and adolescent fertility rate, population density, population growth rate, under-five mortality rate, the life expectancy of males and females, mortality rates of males, risk of maternal death, mortality from road traffic injury, the incidence of tuberculosis, as proxies for overall health, internet penetration rate, and secondary school education attainment levels in the country. We then calculate the pairwise Euclidean distances of each of the 14 standardized variables of these 126 geographical regions with the 51 POLYMOD and DHS countries. We generated 10 000 bootstrap samples of the pairwise distances between countries (between indicators) and selected the combination that maximized the correlation between P_a^c and $\sum_\alpha P_\alpha^c \mathbf{h}_{a,\alpha}^c$. For each of the 126 geographical regions, the reciprocal of the bootstrapped pairwise distances provides the weights which we then use to derive the weighted mean of the population ratio matrices of the 51 POLYMOD and DHS countries.

3. **HAM projection:** Using the derived weighted mean of the population ratio matrices, we project the HAM $\mathbf{h}_{a,\alpha}^c$ for geographical region c.

1.2.3 Household age matrix (HAM) validation for POLYMOD and DHS countries

In section 1.2.2, we describe the methods used to project the HAM of a country.

We perform internal validation using leave-one-out validation to verify that the HAM describing household structure could be reverse-engineered for the POLYMOD and DHS countries for which empirical household

Table 1: Description of Demographic Household Surveys from 43 countries.

Country	Country code*	Survey year	Survey Type	Number of households	Number of individuals
Afghanistan	AFG	2015	Standard DHS	24395	203708
Angola	AGO	2015	Standard DHS	16109	74902
Bangladesh	BGD	2014	Standard DHS	17300	81624
Benin	BEN	2017	Standard DHS	14156	74673
Bolivia (Plurinational State of)	BOL	2008	Standard DHS	19564	77757
Cambodia	KHM	2014	Standard DHS	15825	74122
Cameroon	CMR	2011	Standard DHS	14214	72622
Chad	TCD	2014	Standard DHS	17233	99620
Colombia	COL	2015	Standard DHS	44614	162459
Congo	COG	2011	Standard DHS	11632	51449
Democratic Republic of the Congo	COD	2013	Standard DHS	18171	95949
Dominican Republic	DOM	2013	Standard DHS	11464	41267
Ethiopia	ETH	2016	Standard DHS	16650	75224
Ghana	GHA	2017	Standard DHS	11835	43945
Guatemala	GTM	2014	Standard DHS	21383	102510
Guinea	GIN	2018	Standard DHS	7912	49543
Guyana	GUY	2009	Standard DHS	5632	22845
Haiti	HTI	2016	Standard DHS	13405	59547
Honduras	HND	2011	Standard DHS	21362	100555
India	IND	2015	Standard DHS	601509	2869043
Indonesia	IDN	2017	Standard DHS	47963	197723
Kenya	KEN	2014	Standard DHS	36430	153840
Kyrgyzstan	KGZ	2012	Standard DHS	8040	35805
Lesotho	LSO	2014	Standard DHS	9402	40197
Liberia	LBR	2013	Standard DHS	9333	48219
Malawi	MWI	2015	Standard DHS	26361	120492
Maldives	MDV	2016	Standard DHS	6050	32656
Mali	MLI	2018	Standard DHS	9510	54571
Nepal	NPL	2016	Standard DHS	11040	49064
Niger	NER	2012	Standard DHS	10750	64011
Nigeria	NGA	2018	Standard DHS	40427	188010
Pakistan	PAK	2017	Standard DHS	14540	100869
Peru	PER	2012	Continuous DHS	27218	103211
Philippines	PHL	2017	Standard DHS	27496	120273
Senegal	SEN	2017	Continuous DHS	8380	78950
Sierra Leone	SLE	2013	Standard DHS	12629	75299
South Africa	ZAF	2016	Standard DHS	11083	38850
Timor-Leste	TLS	2016	Standard DHS	11502	61496
Togo	TGO	2013	Standard DHS	9549	46577
Uganda	UGA	2016	Standard DHS	19588	91167
Viet Nam	VNM	2005	Standard AIS	6337	26833
Zambia	ZMB	2018	Standard DHS	12831	65454
Zimbabwe	ZWE	2015	Standard DHS	10534	43706

age matrices were available. The steps involve projecting that country's household age structure as if it were unknown (i.e., following steps 1–3 described in section 1.2.2) and then comparing against the empirical household data to assess the method's performance.

1.3 Working population

We allow the number of age-specific contacts made at the workplace to depend on the age structure of the workforce. To derive the working population matrices for each geographical location, we used the 2019 labour force participation rate by sex and 5-year age groups w_a for the 177 geographical regions from the International Labour Organization (ILO) [7] to compute a joint distribution of the working population, $\mathbf{W}$. The working population, $\mathbf{W}^c$, a square matrix with elements describing the probability of encounters between two ages—a and α—in the workforce of country/geographical region c, is given by

$$\mathbf{W}^c_{a,\alpha} = w^c_a \times w^c_\alpha.$$

After constructing the working population distribution of ages $\mathbf{W}^c$, we project the age-specific contact patterns in the workplace for non-POLYMOD countries.

1.4 School-going population

Similarly, for age-specific contact patterns at school, we construct the age-specific school-going population, including teachers, before projecting the contact patterns in schools. Information on enrolment rates, age ranges of students, pupil-to-teacher ratios by education levels (i.e. pre-primary, primary, secondary, and tertiary) were obtained from UIS [6] and the distribution of teachers by age from Organisation for Economic Co-operation and Development (OECD) [8].

To project the school population, we perform the following steps:

1. **Number of students estimation:** We first estimate the proportion of students in each age interval using the country-specific enrolment rates and the starting and ending ages of students by education level. Together with the population age structure P^c_a, we then estimate the number of students s^c_a aged a in country c. We also estimate the number of students in each education level l, s^c_l, for the education levels pre-primary, primary, secondary, and tertiary.

2. **Number of teachers estimation:** After estimating the number of students for each education level in country c, s^c_l, we use the country-specific pupil-to-teacher ratio to estimate the number of teachers for each education level, l, t^c_l. Together with labour force participation rates by aged and the distribution of teachers by ages, we then estimate the number of teachers t^c_a aged a in country c.

3. **School-going population construction:** We project the school population distribution of ages, as follows:
$$\mathbf{S}^c_{a,\alpha} = \left(\frac{s^c_a}{\sum_i s^c_i} + \frac{t^c_a}{\sum_i t^c_i} \right) \times \left(\frac{s^c_\alpha}{\sum_i s^c_i} + \frac{t^c_\alpha}{\sum_i t^c_i} \right).$$
The elements of $\mathbf{S}^c_{a,\alpha}$ present the probability of encounters between two ages within schools.

1.5 Age- and location-specifc contacts

In Prem et al. (2017) [5], we employed a Bayesian hierarchical modelling framework to estimate modelling to estimate the proclivity of age- and location-specific contact patterns in the POLYMOD countries. The framework estimates both individual-level and population-level parameters and addresses the multi-level structure of the data, accounting for repeat measurements of contacts made in different settings by the same individual in the data. The parameter $\lambda^L_{a,\alpha}$ is the key estimand in the model, and it quantifies typical contact rates between individuals of age groups a and α at location L. With the projected populations at home, work

and school available for the 177 geographical regions (listed in Table 2), we can now deduce the possible age- and location-specific contact matrices:

- **Age-specific contacts at home:** Given the projected HAM $\mathbf{h}_{a,\alpha}^c$ for country c, we can now deduce the possible age-specific contact patterns at home for country c to be:

$$\mu_{a,\alpha}^{H,c} = \lambda_{a,\alpha}^H \times \mathbf{h}_{a,\alpha}^c.$$

- **Age-specific contacts in the workplace:** Having projected the age-specific working populations for the 177 geographical regions $\mathbf{W}_{a,\alpha}^c$, we can infer the age-specific contact in the workplace for country c by the following expression:

$$\mu_{a,\alpha}^{W,c} = \lambda_{a,\alpha}^W \times \mathbf{W}_{a,\alpha}^c \times \frac{P_\alpha^c}{P_\alpha^{POLYMOD}}.$$

- **Age-specific contact in school:** With the projected school-going population available for the 177 geographical regions $\mathbf{S}_{a,\alpha}^c$, we can now deduce the possible age-specific contact patterns in school for country c, as follows:

$$\mu_{a,\alpha}^{S,c} = \lambda_{a,\alpha}^S \times \mathbf{S}_{a,\alpha}^c \times \frac{P_\alpha^c}{P_\alpha^{POLYMOD}}.$$

- **Age-specific contact ar other locations:** We estimate the possible age-specific contacts at other locations (i.e., non-home, work, or school) for country c, as follows:

$$\mu_{a,\alpha}^{O,c} = \lambda_{a,\alpha}^O \times \frac{P_\alpha^c}{P_\alpha^{POLYMOD}}.$$

Table 2: Geographical regions included in the study

Regions	County code	Regions	County code
Afghanistan	AFG	Ecuador	ECU
Albania	ALB	Egypt	EGY
Algeria	DZA	El Salvador	SLV
Angola	AGO	Equatorial Guinea	GNQ
Argentina	ARG	Eritrea	ERI
Armenia	ARM	Estonia	EST
Austria	AUT	Eswatini	SWZ
Azerbaijan	AZE	Ethiopia	ETH
Bahamas	BHS	Fiji	FJI
Bahrain	BHR	Finland	FIN
Bangladesh	BGD	France	FRA
Barbados	BRB	Gabon	GAB
Belarus	BLR	Gambia	GMB
Belgium	BEL	Georgia	GEO
Belize	BLZ	Germany	DEU
Benin	BEN	Ghana	GHA
Bhutan	BTN	Greece	GRC
Bolivia (Plurinational State of)	BOL	Guatemala	GTM
Bosnia and Herzegovina	BIH	Guinea	GIN
Botswana	BWA	Guinea-Bissau	GNB
Brazil	BRA	Guyana	GUY
Brunei Darussalam	BRN	Honduras	HND
Bulgaria	BGR	Hungary	HUN
Burkina Faso	BFA	Iceland	ISL
Burundi	BDI	India	IND
Cabo Verde	CPV	Indonesia	IDN
Cambodia	KHM	Iran (Islamic Republic of)	IRN
Cameroon	CMR	Iraq	IRQ
Canada	CAN	Ireland	IRL
Central African Republic	CAF	Israel	ISR
Chad	TCD	Italy	ITA
Chile	CHL	Jamaica	JAM
China	CHN	Jordan	JOR
China, Hong Kong SAR	HKG	Kazakhstan	KAZ
China, Macao SAR	MAC	Kenya	KEN
Colombia	COL	Kuwait	KWT
Comoros	COM	Kyrgyzstan	KGZ
Congo	COG	Lao People's Democratic Republic	LAO
Costa Rica	CRI	Latvia	LVA
Côte d'Ivoire	CIV	Lesotho	LSO
Croatia	HRV	Liberia	LBR
Cuba	CUB	Libya	LBY
Cyprus	CYP	Lithuania	LTU
Czechia	CZE	Luxembourg	LUX
Dem. People's Republic of Korea	PRK	Madagascar	MDG
Democratic Republic of the Congo	COD	Malawi	MWI
Denmark	DNK	Malaysia	MYS
Djibouti	DJI	Maldives	MDV
Dominican Republic	DOM	Mali	MLI

Regions	County code	Regions	County code
Malta	MLT	Sri Lanka	LKA
Mauritania	MRT	State of Palestine	PSE
Mauritius	MUS	Sudan	SDN
Mexico	MEX	Suriname	SUR
Mongolia	MNG	Sweden	SWE
Montenegro	MNE	Switzerland	CHE
Morocco	MAR	Syrian Arab Republic	SYR
Mozambique	MOZ	Tajikistan	TJK
Myanmar	MMR	Thailand	THA
Namibia	NAM	Timor-Leste	TLS
Nepal	NPL	Togo	TGO
Netherlands	NLD	Tonga	TON
New Zealand	NZL	Trinidad and Tobago	TTO
Nicaragua	NIC	Tunisia	TUN
Niger	NER	Turkey	TUR
Nigeria	NGA	Turkmenistan	TKM
North Macedonia	MKD	Uganda	UGA
Norway	NOR	Ukraine	UKR
Oman	OMN	United Arab Emirates	ARE
Pakistan	PAK	United Kingdom	GBR
Panama	PAN	United Republic of Tanzania	TZA
Papua New Guinea	PNG	United States of America	USA
Paraguay	PRY	Uruguay	URY
Peru	PER	Uzbekistan	UZB
Philippines	PHL	Vanuatu	VUT
Poland	POL	Venezuela (Bolivarian Republic of)	VEN
Portugal	PRT	Viet Nam	VNM
Puerto Rico	PRI	Yemen	YEM
Qatar	QAT	Zambia	ZMB
Republic of Korea	KOR	Zimbabwe	ZWE
Republic of Moldova	MDA		
Romania	ROU		
Russian Federation	RUS		
Rwanda	RWA		
Saint Lucia	LCA		
Saint Vincent and the Grenadines	VCT		
Samoa	WSM		
Sao Tome and Principe	STP		
Saudi Arabia	SAU		
Senegal	SEN		
Serbia	SRB		
Sierra Leone	SLE		
Singapore	SGP		
Slovakia	SVK		
Slovenia	SVN		
Solomon Islands	SLB		
South Africa	ZAF		
South Sudan	SSD		
Spain	ESP		

1.6 Age- and location-specifc contacts by rural and urban areas

The population age structure at the various locations—in the household, workplaces, schools, general community—varies by rural and urban subregions of country. Hence, we would expect the contact patterns within a country to also vary by rural and urban subregions. To account for these differences, we also stratified the age- and location-specific contact matrices by rural-urban areas.

The United Nations Population Division provides the estimation of rural and urban population age compositions for the 177 geographical regions [2]. The nationally-representative DHS household surveys provide data for rural and urban areas, allowing us to derive rural-urban HAM and project rural-urban HAM for the countries with no available household data (following steps 1–3 described in section 1.2.2). ILO providesthe age-specific labour force participation rates by rural and urban regions [9], allowing us to determine the working population age structure, $\mathbf{W}_{a,\alpha}^c$, for country c by rural-urban subregions. Using the differences in rural and urban schools' pupil-to-teacher from the OECD [10], we construct rural and urban school-going population age structures, $\mathbf{S}_{a,\alpha}^c$.

After projecting the populations at home, work and school for rural and urban subregions, we can then deduce the possible age- and location-specific contact matrices following the steps detailed in section 1.5.

1.7 Age-stratified compartmental model of the physical distancing interventions for COVID-19

We adapt a discrete-time Susceptible-Exposed-Infectious-Recovered (SEIR) compartmental model [11,12] presented in Figure 1. The model population is closed, and it ignores all demographic changes in the population (i.e., births, deaths, and ageing). The model stratifies the population into (i) disease states: susceptible, exposed (infected but not yet infectious), infectious (i.e., preclinical, clinical, or subclinical), and recovered (or died/removed) states; and (ii) 5-year age groups until age 70 years and a single category aged ≥ 75 years, resulting in 16 age bands.

We implement the model stochastically and allow it to select random values of parameters from the set of uncertainty distributions presented in Table 3, adapted from [11,12].
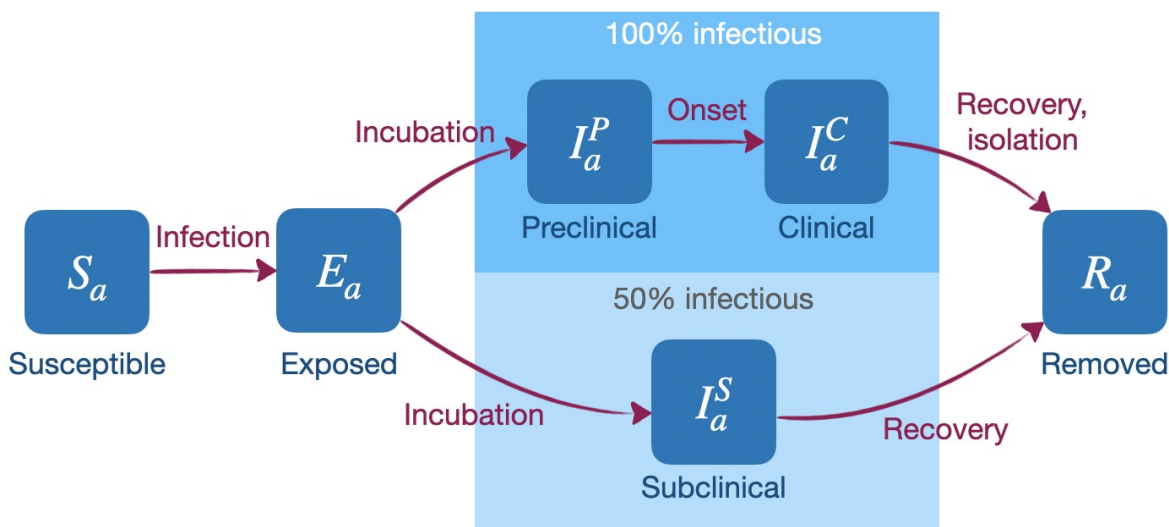
Figure 1: Age-stratified stochastic SEIR model. Adapted from Davies et al. (2020) [11].

As described in [11,12], susceptible individuals might acquire the infection when they come in contact with an

Table 3: Parameters of the age-stratified SEIR model.

	Description	Values	References
Δt	Time step for simulation	0.25 days or 6 hours	
d_E	Latent period in days	$d_E \sim \mathrm{Gamma}(4, 4)$	[14–16]
	Duration of infectiousness in days		
d_P	(i) during the preclinical phase	$d_P \sim \mathrm{Gamma}(1.5, 4)$	[17]
d_C	(ii) during the clinical phase	$d_C \sim \mathrm{Gamma}(3.5, 4)$	[14–16]
d_S	(iii) during the subclinical phase	$d_S \sim \mathrm{Gamma}(5.0, 4)$	Assumed to be the same as the total duration of infectiousness for clinical cases (i.e., $d_P + d_C$)
y_a	Probability of clinical symptoms on infection for individuals aged a	Age-dependent, as estimated in [18]	[18]
f	Relative infectiousness of subclinical cases	50%	Assumed
u	Probability of transmission per contact with an infectious individual	Refer to text	Derived
N_a	Number of individuals aged a	Demographic data	[1]
$c_{a,\alpha}$	Number of contacts per day with individuals aged α by an individual age a	Empirical and synthetic contact matrices	
R_0	Basic reproduction number	$R_0 \sim \mathrm{Normal}(2.6, 0.5^2)$	[13]
	Proportion of hospitalised cases requiring critical care	30%	[20]
	Duration of disease in days		
	(i) severe, non-critical disease	$\sim \mathrm{Gamma}(8, 8)$	NHS Digit
	(ii) severe, critical disease	$\sim \mathrm{Gamma}(10, 10)$	[20]
	Delay from symtom onset		
	(i) to becoming a severe case (days)	$\sim \mathrm{Gamma}(7, 7)$	[19,20]
	(i) to death (days)	$\sim \mathrm{Gamma}(22, 22)$	[19,21]

infectious person and enter the exposed disease state before they become infectious and later either recover or die. The time step of the model is 6 hours. We assume age-specific mixing patterns of individuals alter their likelihood of being exposed to the virus given a certain number of infectious people in the population. The force of infection defined as the rate at which susceptible individuals become exposed t, and is given by for any age group a and time increment t as

$$\lambda_{a,t} = u \times \sum_{\alpha} c_{a,\alpha} \times \frac{I_\alpha^P + I_\alpha^C + f \times I_\alpha^S}{N_\alpha}$$

where u is the probability of transmission per contact with an infectious person, $c_{a,\alpha}$ denotes the number of contacts per day with individuals aged α by an individual in age group a per time increment (drawn from the contact matrix), and $\frac{(I_\alpha^P + I_\alpha^C + f I_\alpha^S)}{N_\alpha}$ is the probability that any age α individual contacted is infectious, with f denoting the relative infectiousness of subclinical cases, compared to clinical cases. The basic reproduction number R_0 is defined as the average number of secondary infections generated by an infectious individual in a fully susceptible population, we calculated it as the absolute value of the dominant eigenvalue of the next generation matrix, which was derived by linearising the system at epidemic equilibrium [13]. For any stochastic run, u is derived from the ratio of this eigenvalue and the R_0 value selected for that run. We considered six contact matrices when modelling the interventions to the COVID-19 pandemic: the empirically-constructed contact matrices at the study-year and adjusted for the 2020 population, the 2017 synthetic matrices, and the updated synthetic matrices at the national, rural, or urban settings.

The duration (in days) an individual spends in states E, IP, IC, or IS is drawn from distributions d_E, d_P, d_C and d_S, respectively (Table 3) [14–17]. After the latency period, infected individuals are divided into

clinical (symptomatic) and sub-clinical (asymptomatic) states with probability y_a and $1 - y_a$, respectively [18]. Clinical individuals will first experience a preclinical phase where they are still infectious (denoted as I^P in Figure 1), followed by a clinical and infectious phase (denoted as I^C in Figure 1). We assume that subclinical individuals, represented as I^S in Figure 1, are half as infectious as clinical cases. Clinical and subclinical individuals have the same duration of infectiousness, and we assume that their clinical severity do not affect their infectiousness. These individuals will later recover or die, and we assume that all individuals who have recovered (or have left the infectious phase) are immune until the end of the simulations. We account for delays in the transitions to different clinical disease states (such as symptom onset to severe case, critical or not; and from onset of severe symptoms to recovery or death) [19–21]. These delays are drawn from distributions reported in the literature; however, they do not change the force of infection or transmissibility.

2   Results

2.1   Household age matrix (HAM) validation for POLYMOD and DHS countries

The empirical household age structures (i.e., from the data) and the modelled household age structures (i.e., the leave-one-out validation, assuming no available household data) for each POLYMOD and DHS countries are represented in the first and second panels respectively of the figures in this section. In the third panel, a scatter plot of the entries in the observed (x-axis) and modelled (y-axis) HAM. Using the methods described in section 1.2.2, we reconstructed the empirical household age structures for the POLYMOD and DHS countries with high fidelity (median correlation between the observed and modelled HAM 0.92, with an interquartile range 0.85–0.95).

HAM of Belgium

HAM of Germany

HAM of Finland

HAM of United Kingdom

Mean number of household members in a specific age group

HAM of Italy

Mean number of household members in a specific age group

HAM of Luxembourg

Mean number of household members in a specific age group

HAM of Netherlands

(a) Empirical POLYMOD / DHS data
(b) Modelled HAM
(c) r = 0.8

HAM of Poland

(a) Empirical POLYMOD / DHS data
(b) Modelled HAM
(c) r = 0.54

HAM of Afghanistan

(a) Empirical POLYMOD / DHS data
(b) Modelled HAM
(c) r = 0.92

HAM of Angola

HAM of Bangladesh

HAM of Benin

HAM of Bolivia (Plurinational State of)

HAM of Cambodia

HAM of Cameroon

HAM of Chad

HAM of Colombia

HAM of Congo

HAM of Democratic Republic of the Congo

HAM of Dominican Republic

HAM of Ethiopia

17

HAM of Ghana

HAM of Guatemala

HAM of Guinea

HAM of Guyana

HAM of Haiti

HAM of Honduras

HAM of India

HAM of Indonesia

HAM of Kenya

20

HAM of Kyrgyzstan

Empirical POLYMOD / DHS data

Modelled HAM

Mean number of household members
in a specific age group

(a)

(b)

(c) r = 0.86

HAM of Lesotho

Empirical POLYMOD / DHS data

Modelled HAM

Mean number of household members
in a specific age group

(a)

(b)

(c) r = 0.93

HAM of Liberia

Empirical POLYMOD / DHS data

Modelled HAM

Mean number of household members
in a specific age group

(a)

(b)

(c) r = 0.97

HAM of Malawi

HAM of Maldives

HAM of Mali

22

HAM of Nepal

HAM of Niger

HAM of Nigeria

HAM of Pakistan

HAM of Peru

HAM of Philippines

HAM of Senegal

Empirical POLYMOD / DHS data

Modelled HAM

Mean number of household members in a specific age group

(a)

(b)

(c) r = 0.9

Age of household member

Age of household member

Modelled HAM

Empirical POLYMOD / DHS data

75–79
70–74
65–69
60–64
55–59
50–54
45–49
40–44
35–39
30–34
25–29
20–24
15–19
10–14
5–9
0–4

HAM of Sierra Leone

Empirical POLYMOD / DHS data

Modelled HAM

Mean number of household members in a specific age group

(a)

(b)

(c) r = 0.96

Age of household member

Age of household member

Modelled HAM

Empirical POLYMOD / DHS data

75–79
70–74
65–69
60–64
55–59
50–54
45–49
40–44
35–39
30–34
25–29
20–24
15–19
10–14
5–9
0–4

HAM of South Africa

Empirical POLYMOD / DHS data

Modelled HAM

Mean number of household members in a specific age group

(a)

(b)

(c) r = 0.88

Age of household member

Age of household member

Modelled HAM

Empirical POLYMOD / DHS data

75–79
70–74
65–69
60–64
55–59
50–54
45–49
40–44
35–39
30–34
25–29
20–24
15–19
10–14
5–9
0–4

HAM of Timor−Leste

Empirical POLYMOD / DHS data

Modelled HAM

Mean number of household members
in a specific age group

r = 0.93

HAM of Togo

Empirical POLYMOD / DHS data

Modelled HAM

Mean number of household members
in a specific age group

r = 0.98

HAM of Uganda

Empirical POLYMOD / DHS data

Modelled HAM

Mean number of household members
in a specific age group

r = 0.92

HAM of Viet Nam

(a) Empirical POLYMOD / DHS data

(b) Modelled HAM

(c) r = 0.75

HAM of Zambia

(a) Empirical POLYMOD / DHS data

(b) Modelled HAM

(c) r = 0.94

HAM of Zimbabwe

(a) Empirical POLYMOD / DHS data

(b) Modelled HAM

(c) r = 0.93

2.2 Household age matrix (HAM) urban and rural comparison

Using rural-urban stratified population and household data, we present the population age compositions and household age matrices for rural and urban areas by the rural and urban subregions for countries with recent DHS household surveys.

Empirical DHS household data: Afghanistan

Empirical DHS household data: Angola
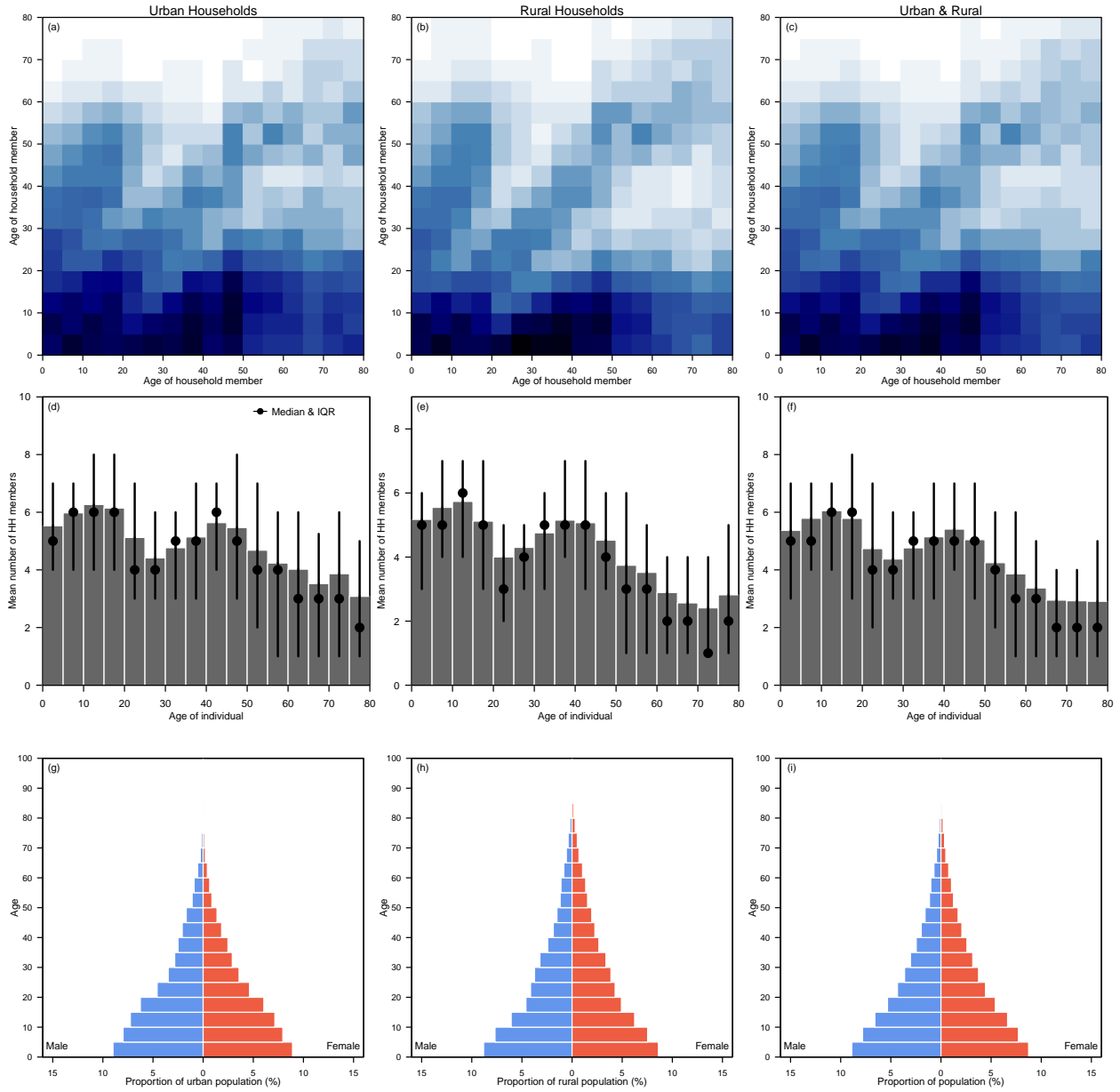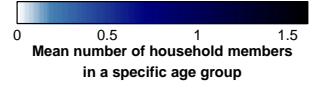
Empirical DHS household data: Bangladesh
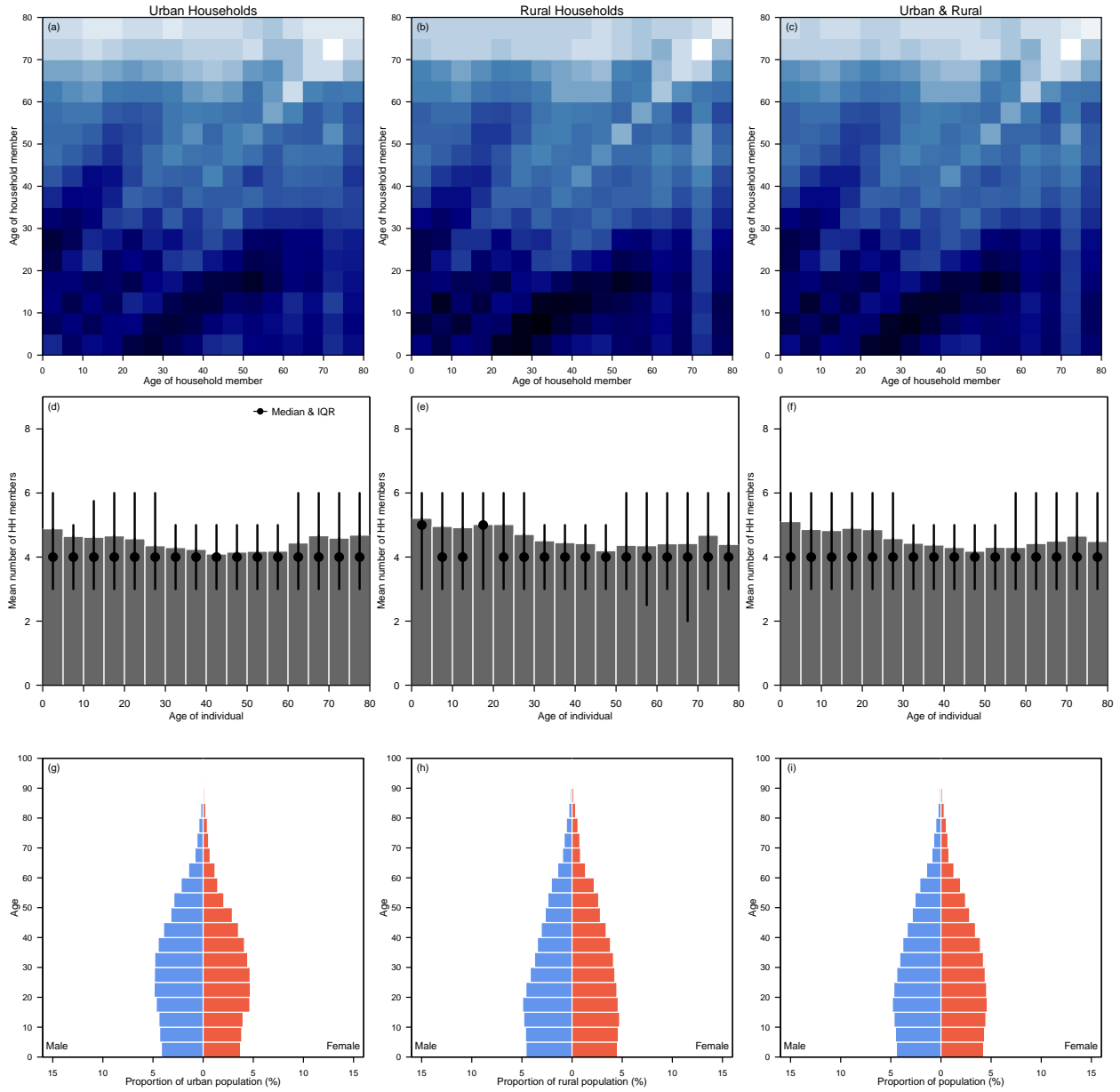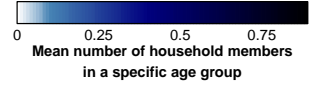
Empirical DHS household data: Benin
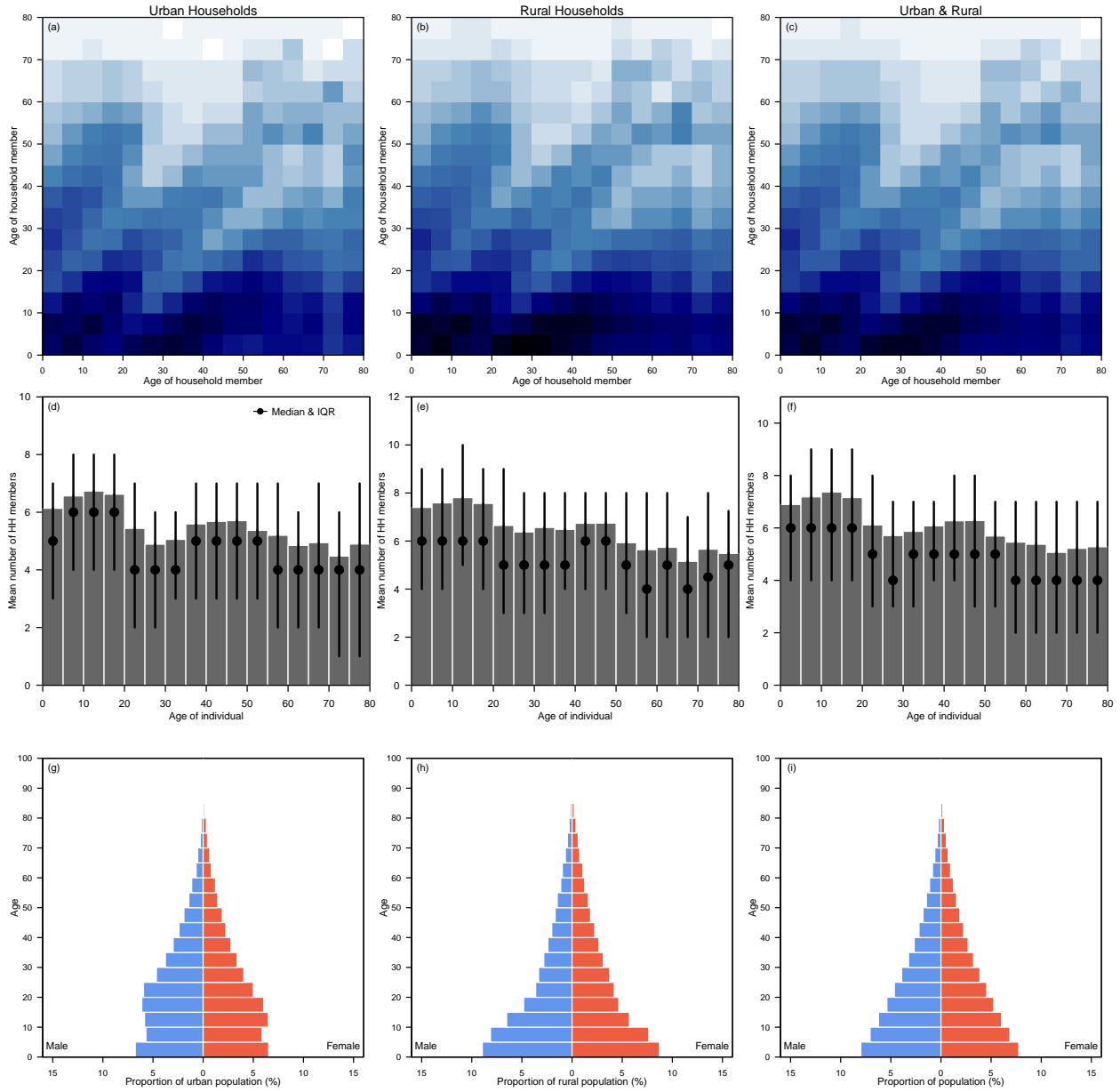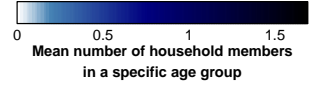
Empirical DHS household data: Bolivia (Plurinational State of)
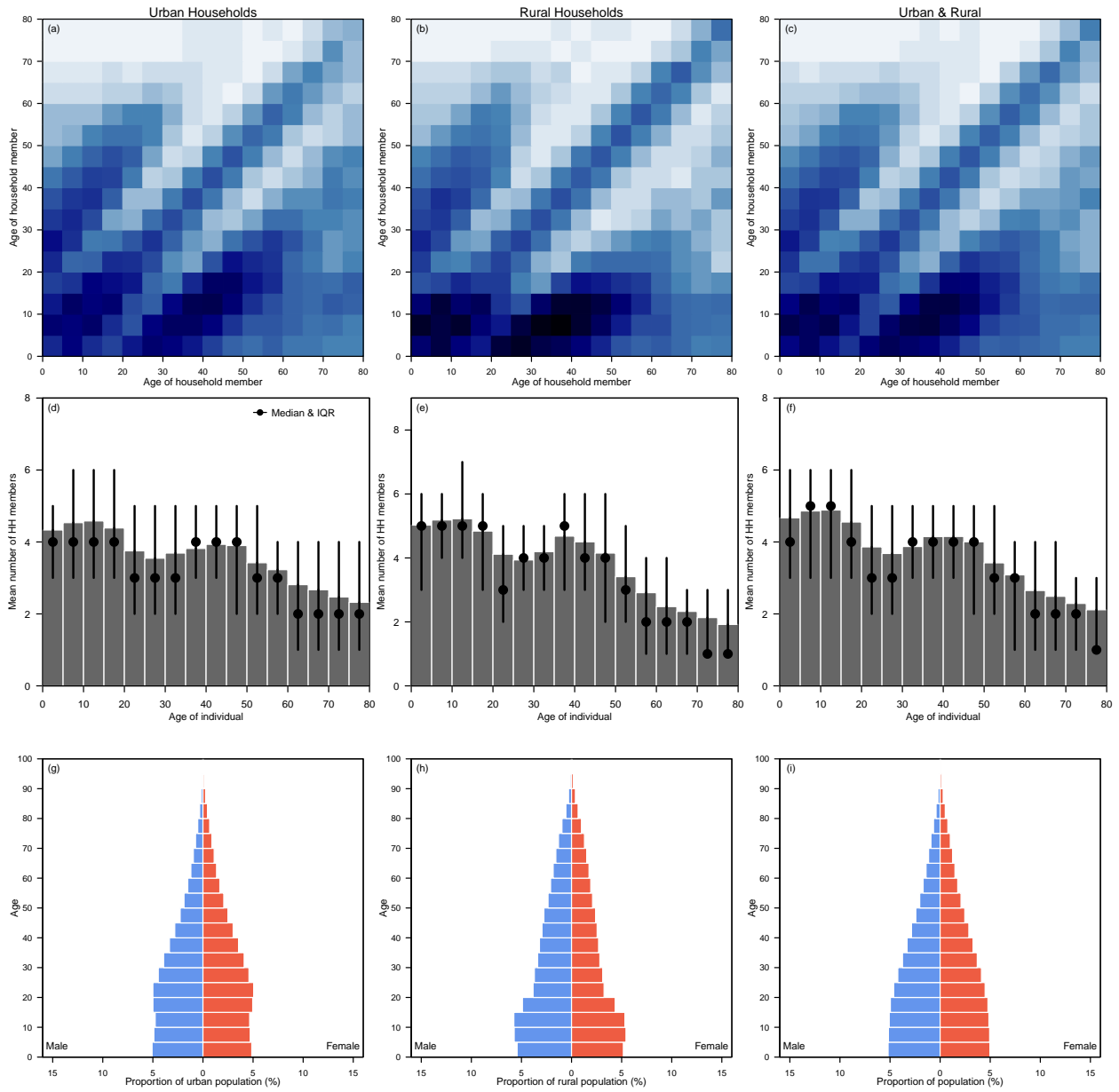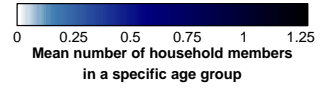
Empirical DHS household data: Cambodia

Mean number of household members
in a specific age group

Empirical DHS household data: Cameroon

Mean number of household members in a specific age group
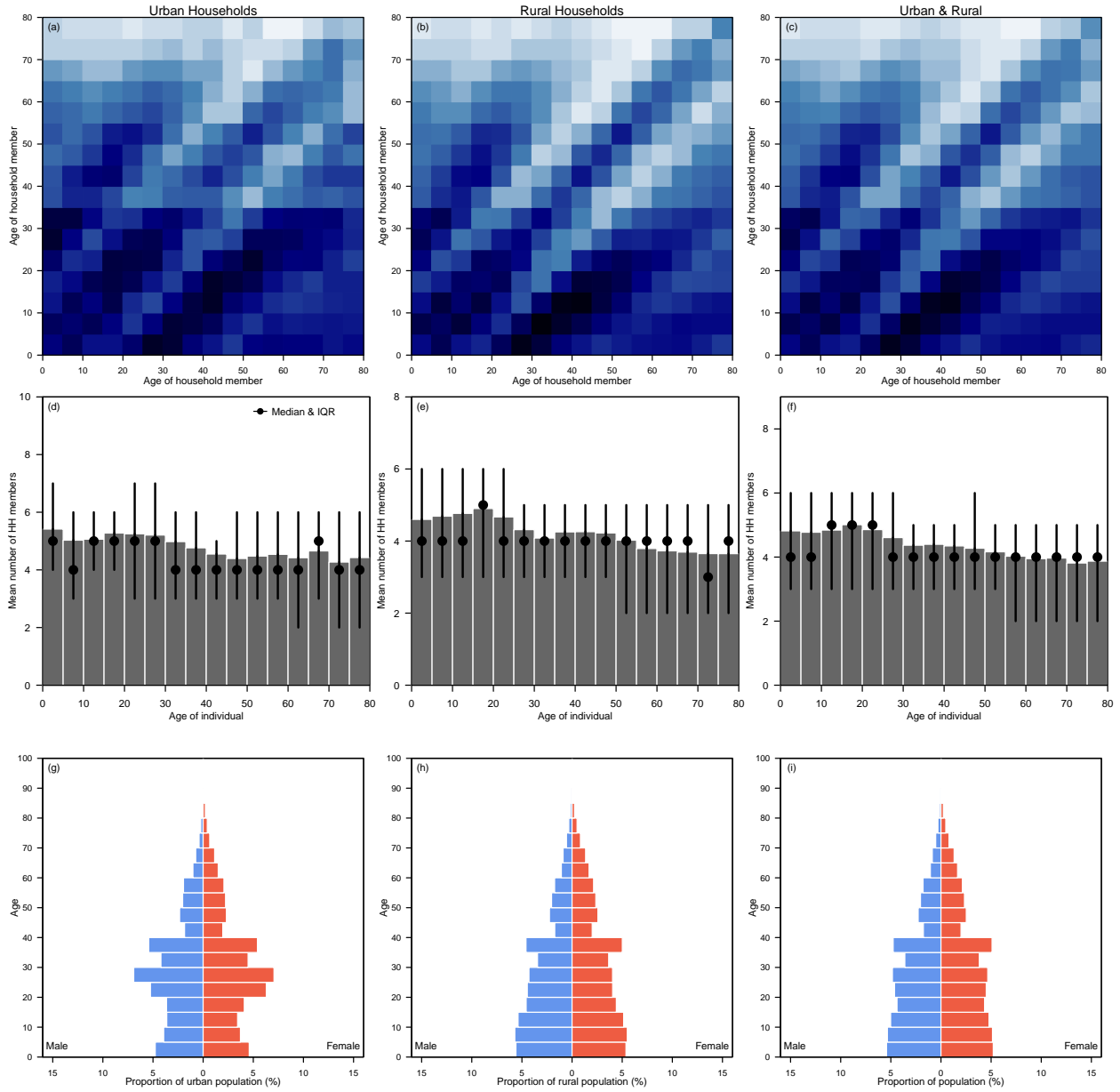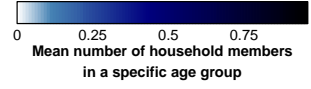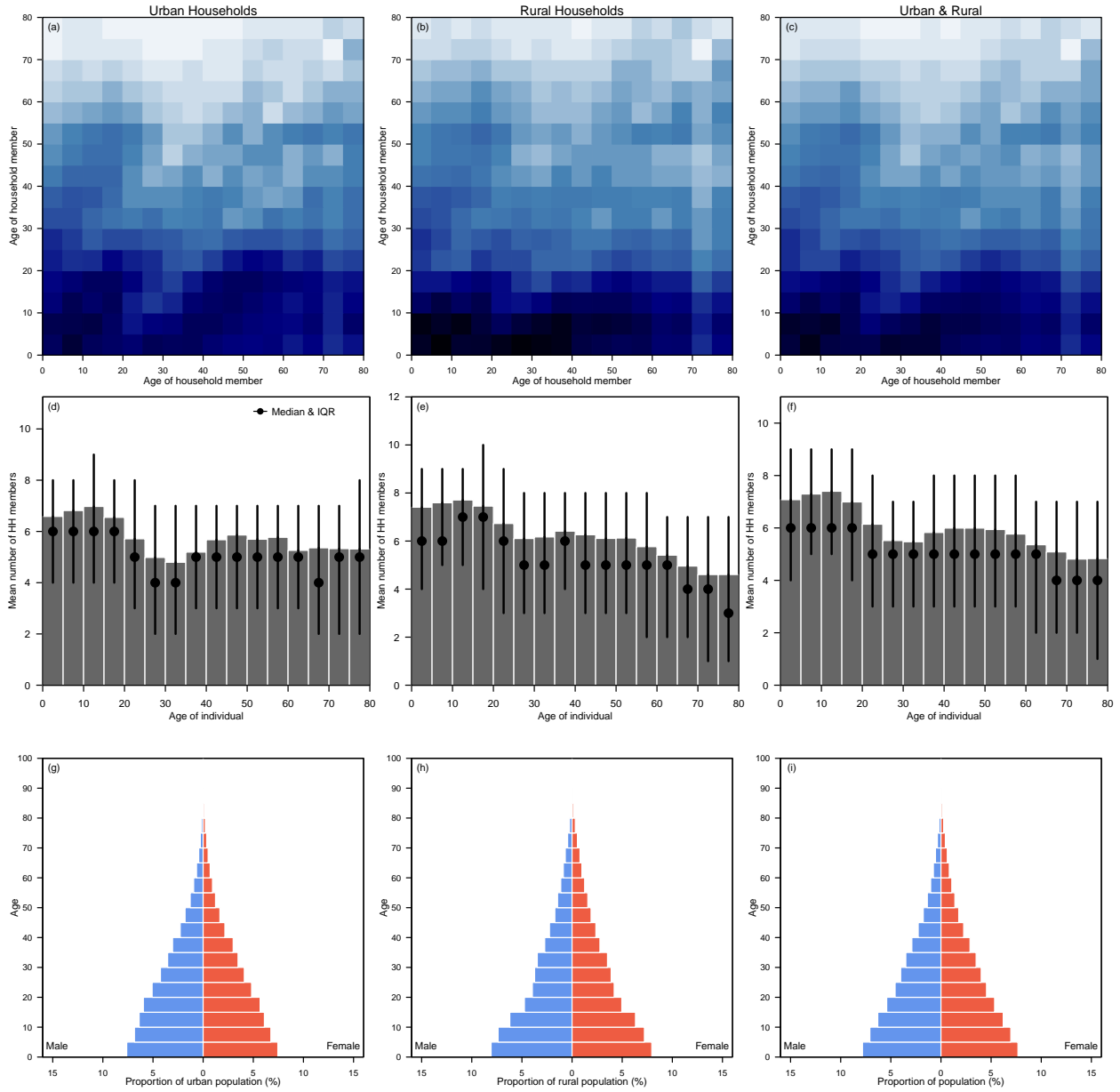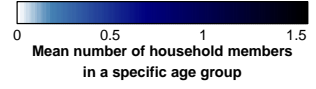
Empirical DHS household data: Chad
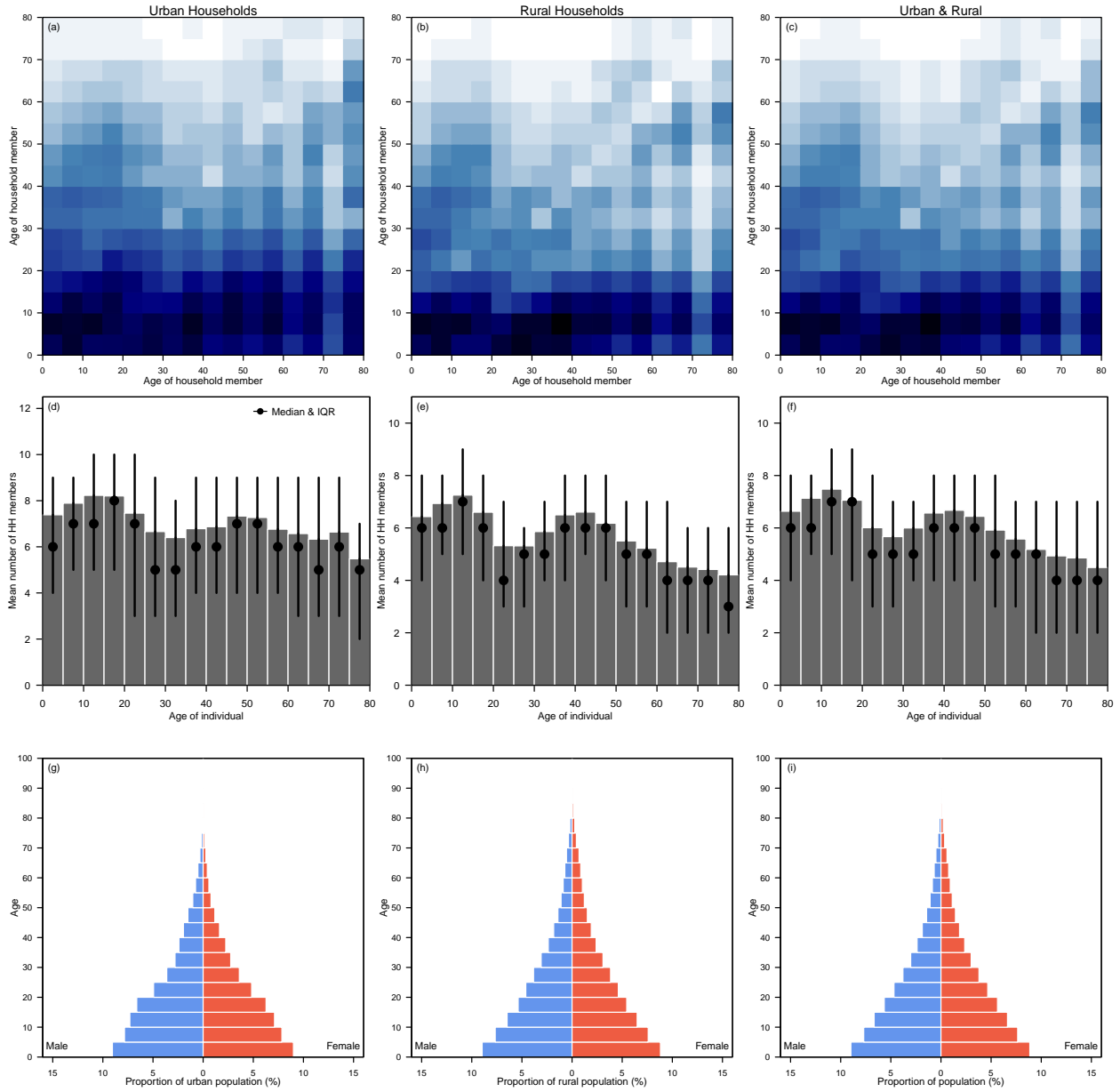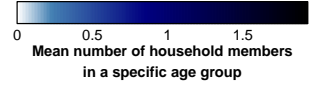
Empirical DHS household data: Colombia

Mean number of household members
in a specific age group
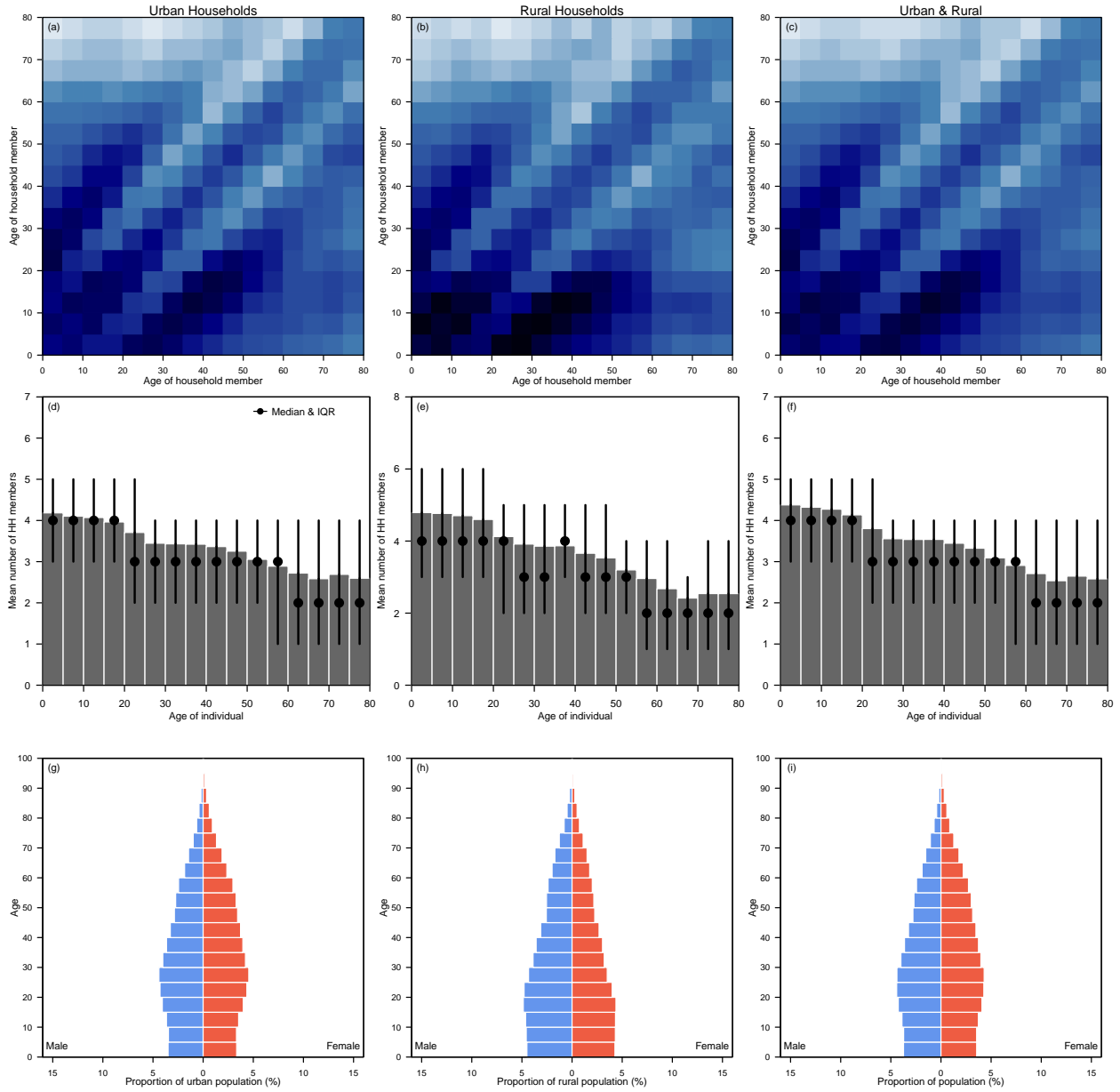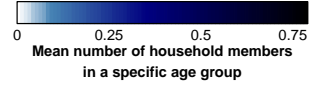
Empirical DHS household data: Congo
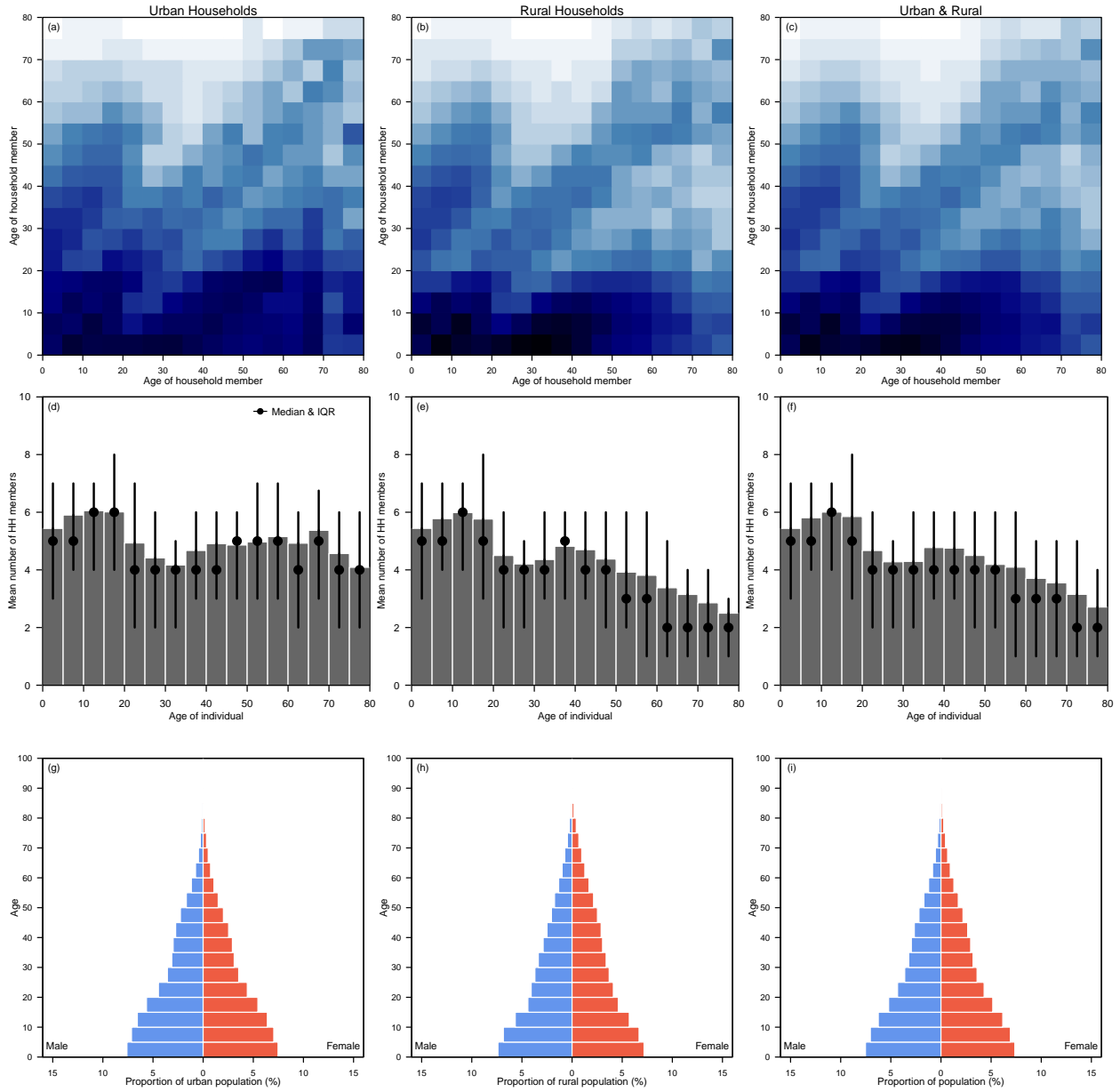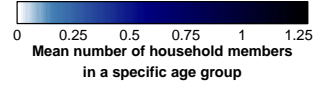
Empirical DHS household data: Democratic Republic of the Congo

Empirical DHS household data: Dominican Republic

Mean number of household members
in a specific age group

Urban Households

Rural Households

Urban & Rural
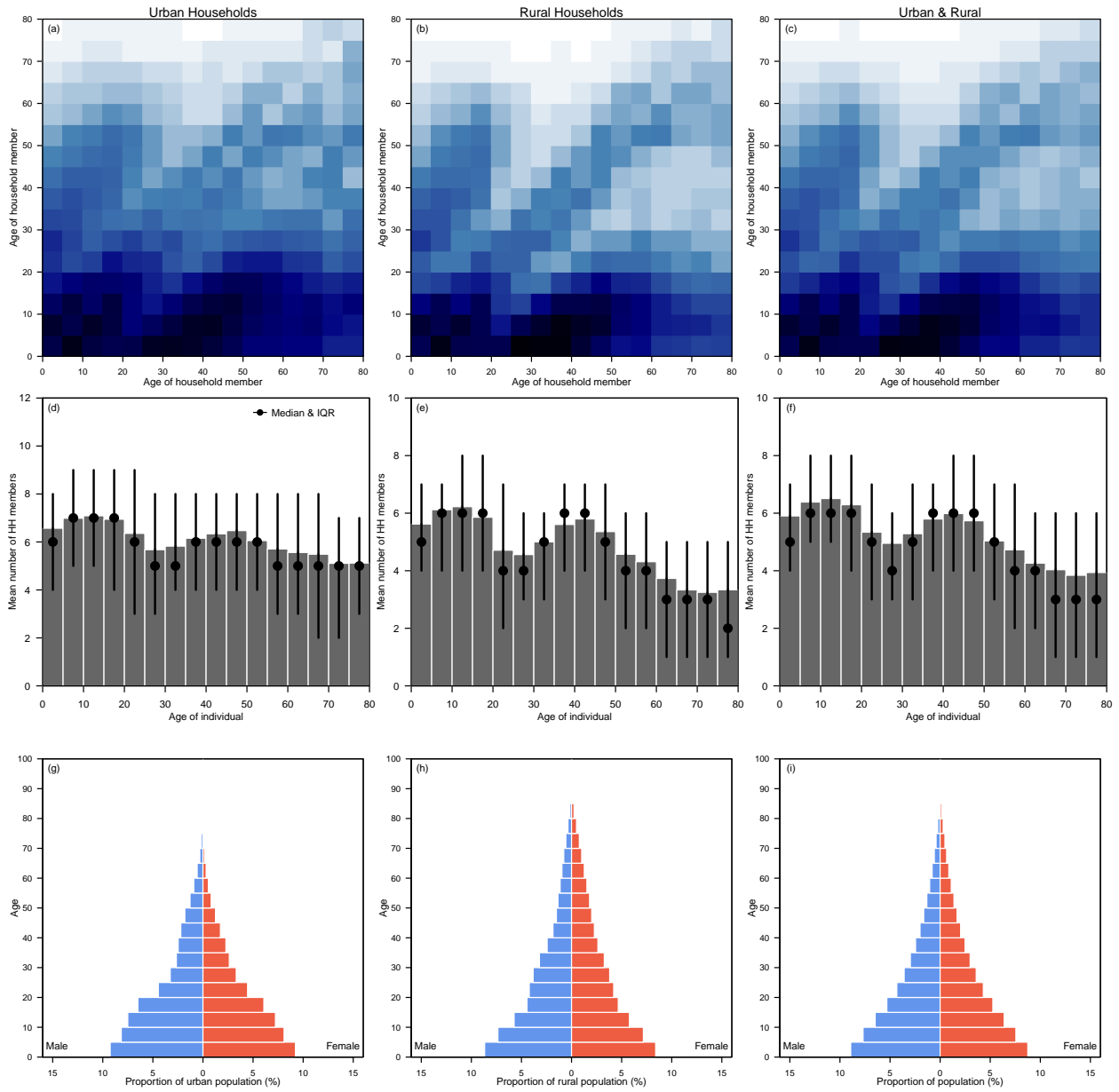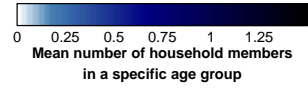
Empirical DHS household data: Ethiopia

Empirical DHS household data: Ghana

Mean number of household members in a specific age group

Urban Households

Rural Households

Urban & Rural
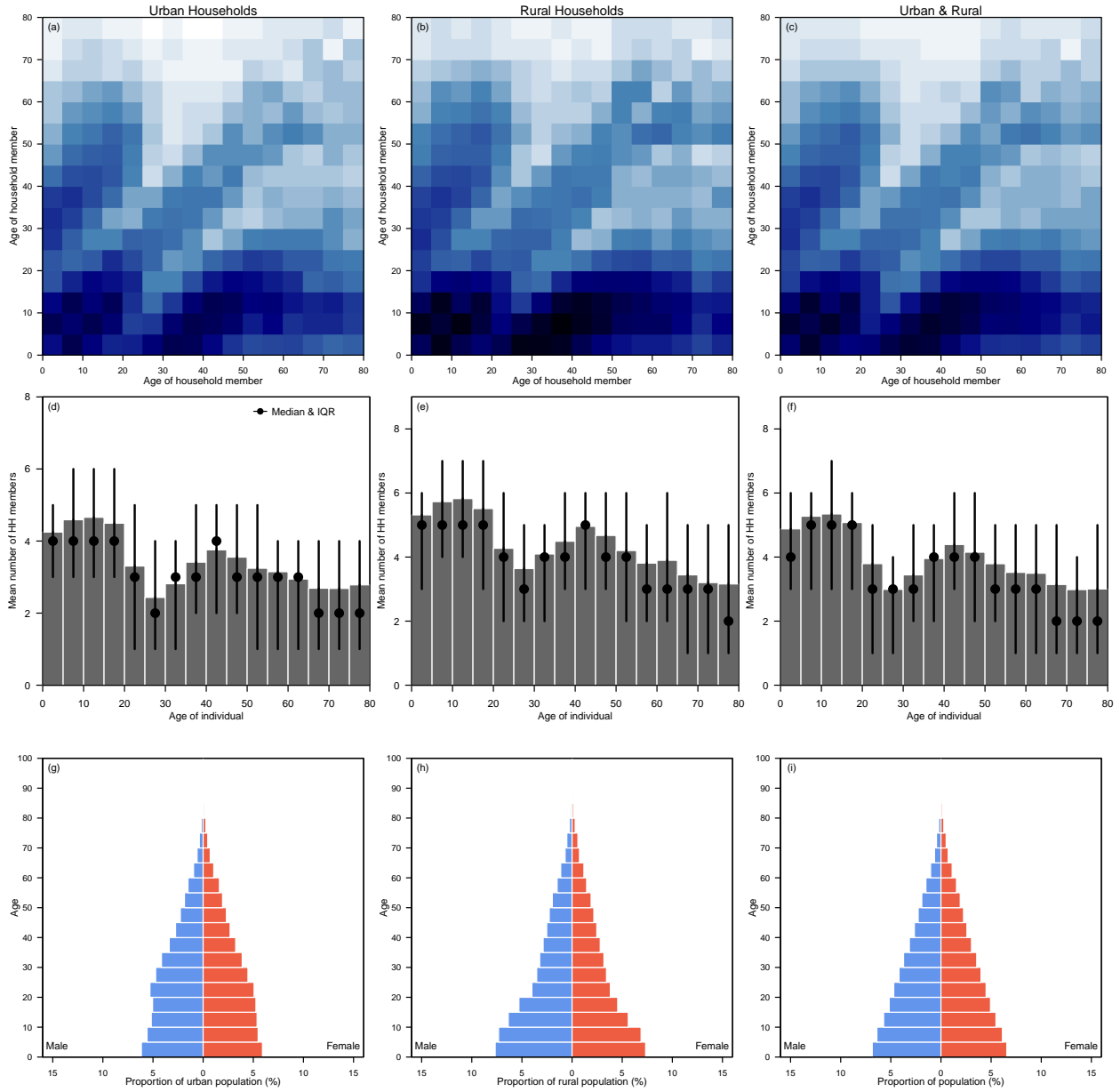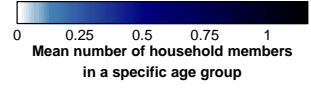
Empirical DHS household data: Guatemala

Mean number of household members in a specific age group

Urban Households
Rural Households
Urban & Rural
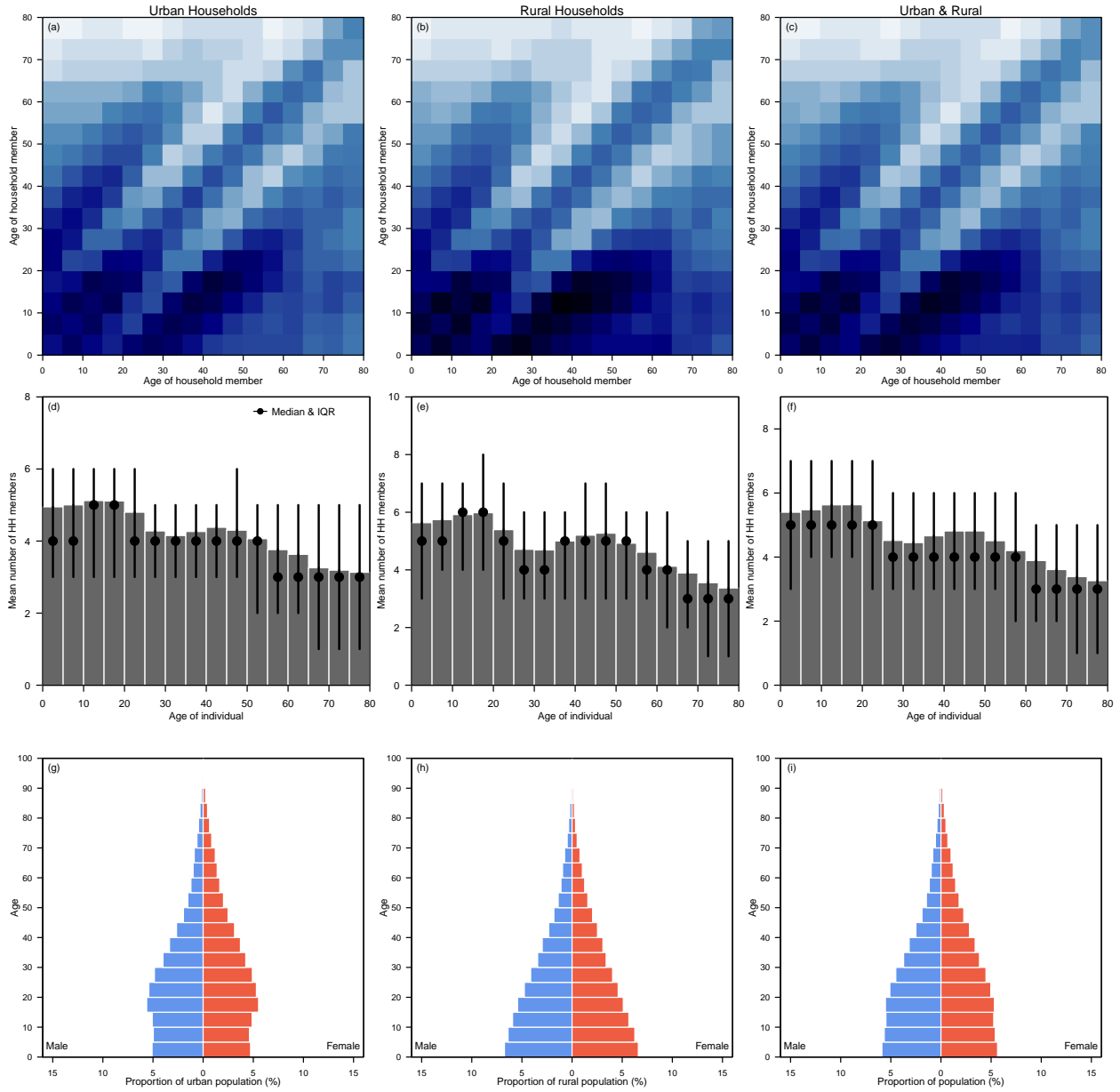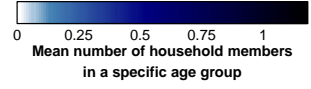
Empirical DHS household data: Guinea
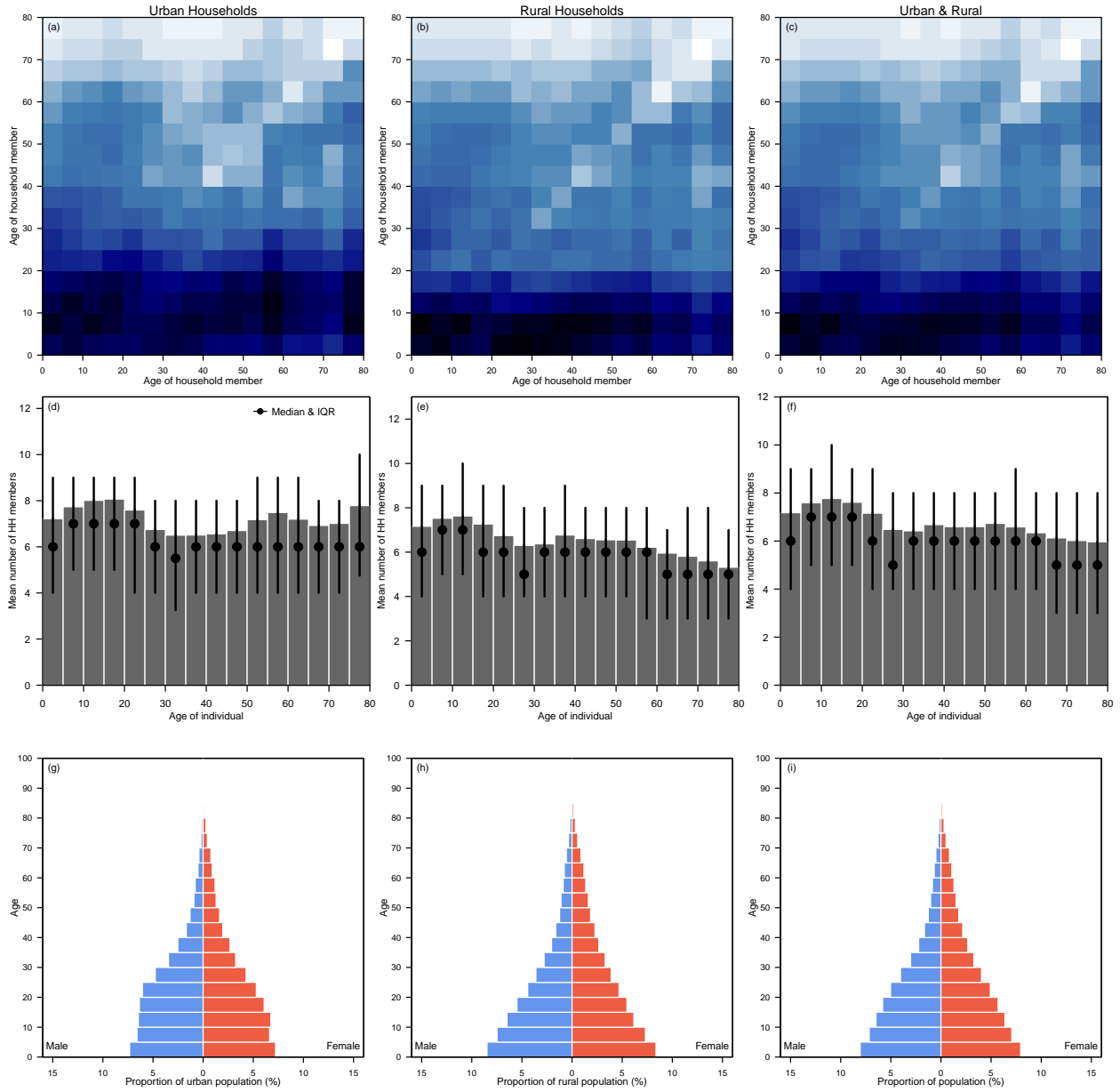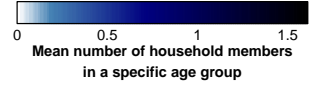
Empirical DHS household data: Guyana

Empirical DHS household data: Haiti

45

Empirical DHS household data: Honduras

Mean number of household members in a specific age group
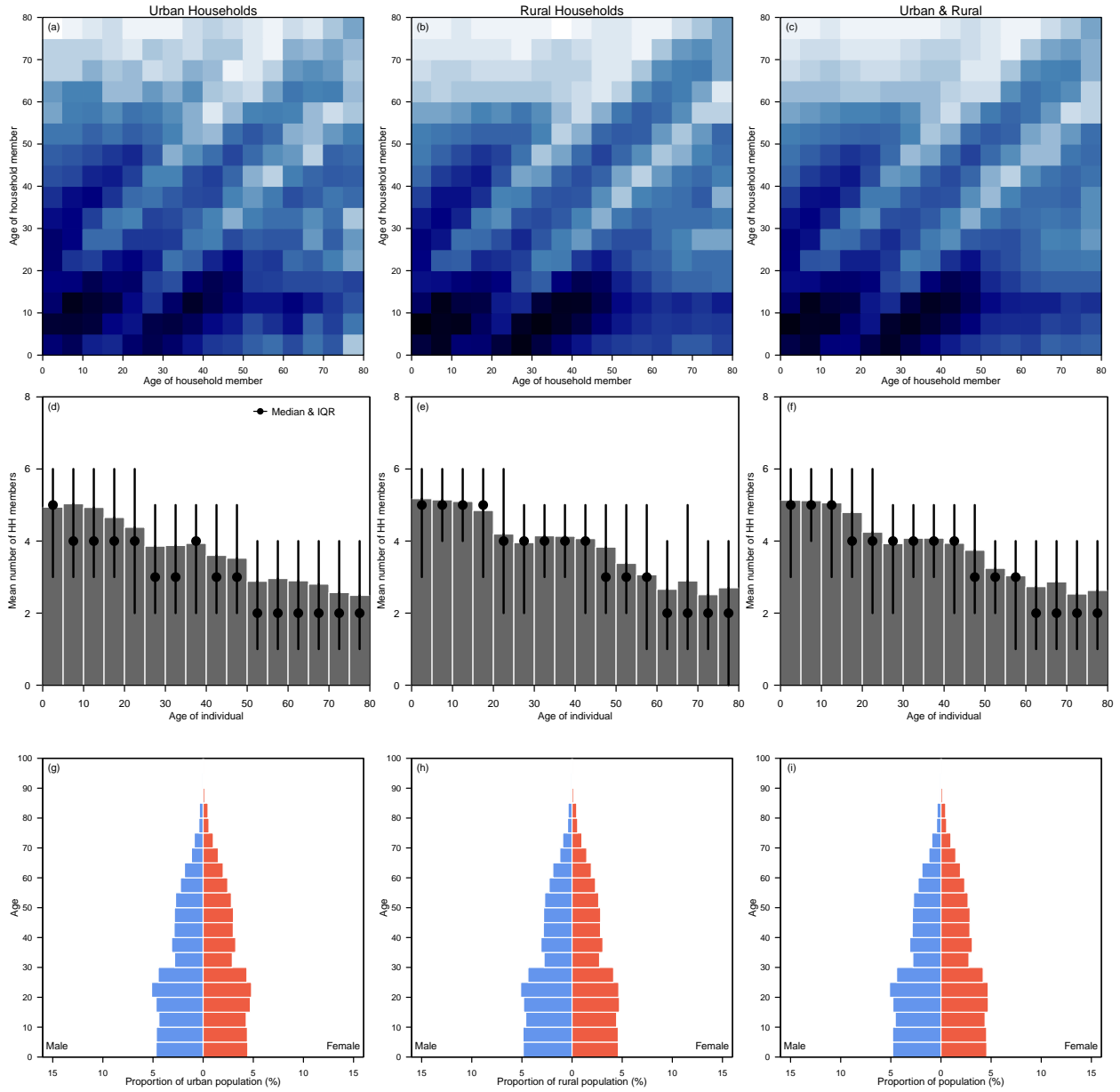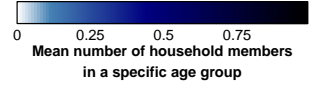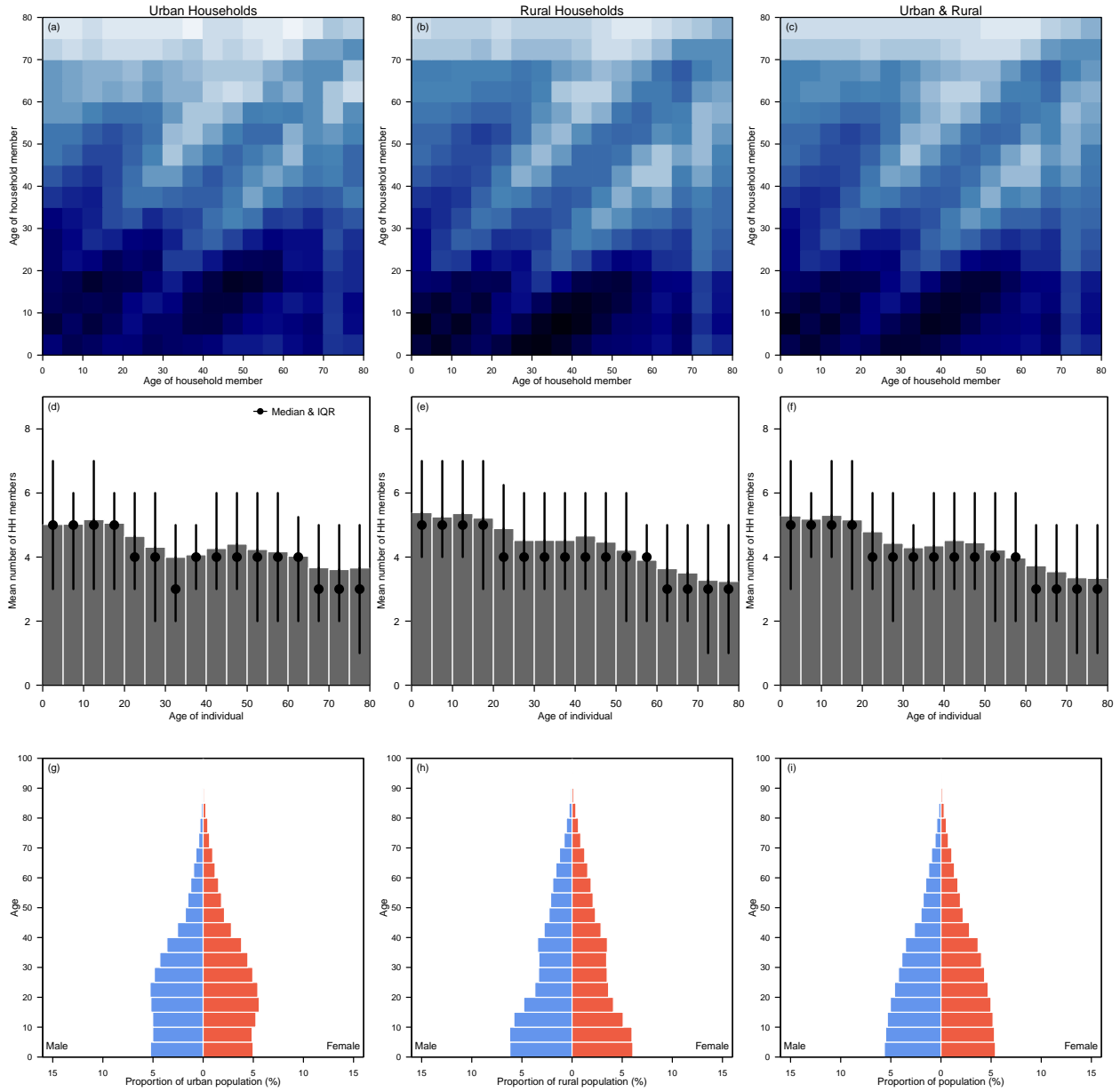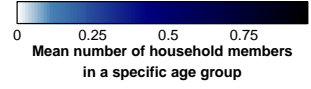
Empirical DHS household data: Honduras
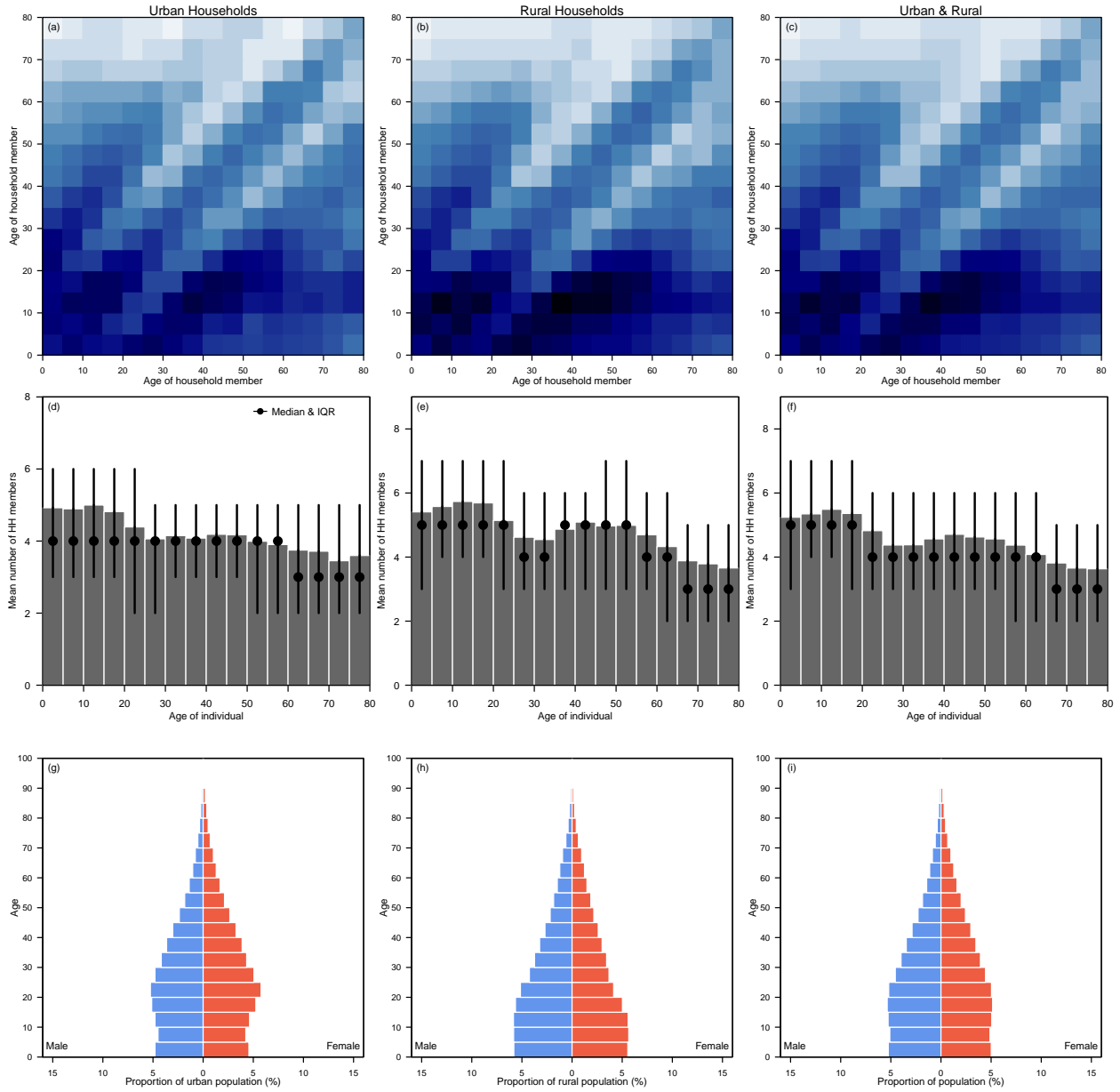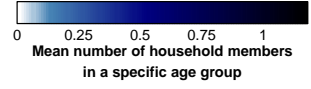
Empirical DHS household data: India

Empirical DHS household data: Indonesia

Empirical DHS household data: Kenya

Mean number of household members in a specific age group

0   0.25   0.5   0.75   1

Urban Households

(a)

Rural Households

(b)

Urban & Rural

(c)

(d)   ● Median & IQR

(e)

(f)

(g)   Male   Female

(h)   Male   Female

(i)   Male   Female
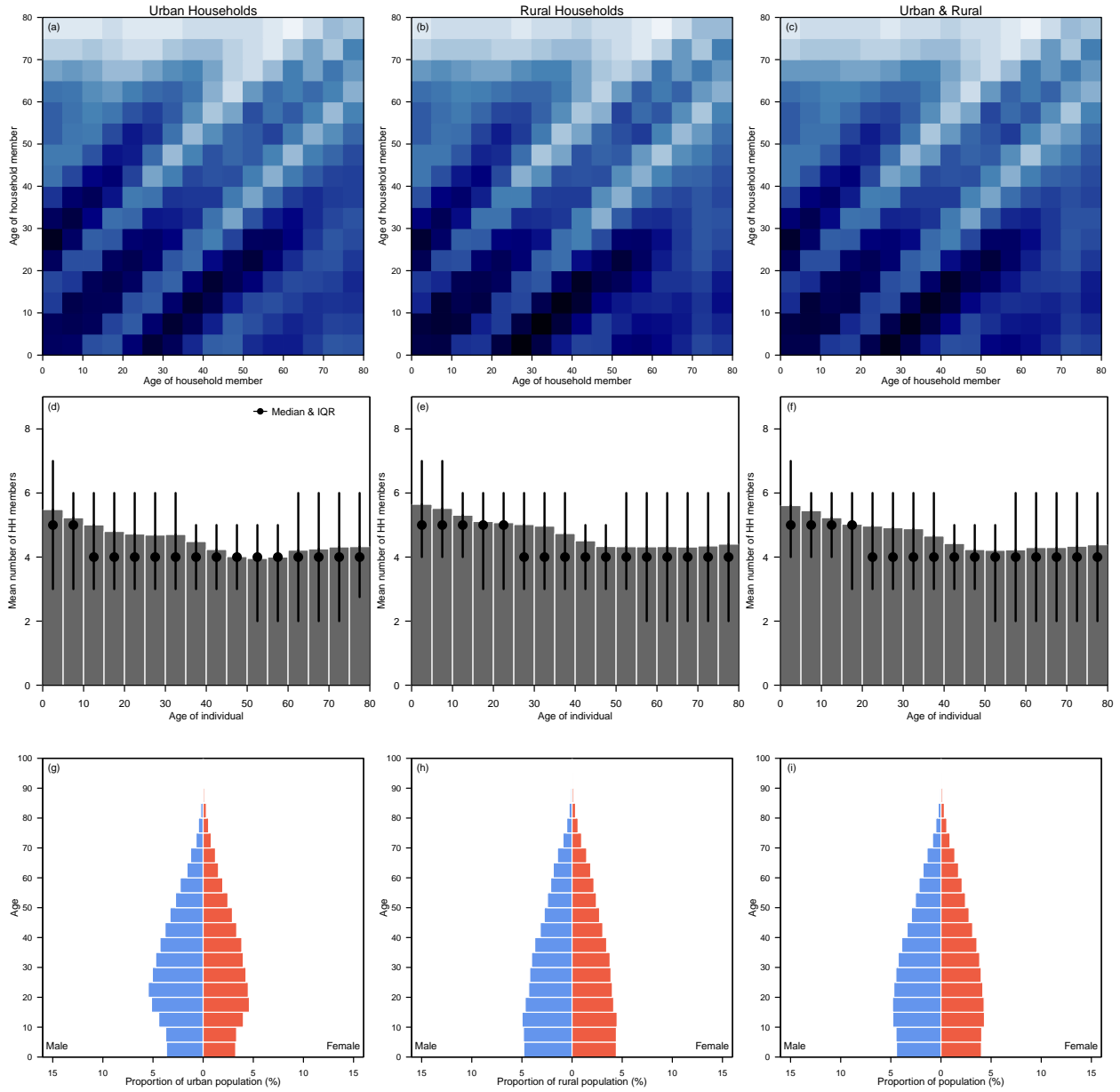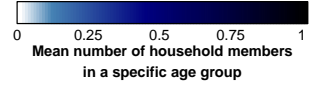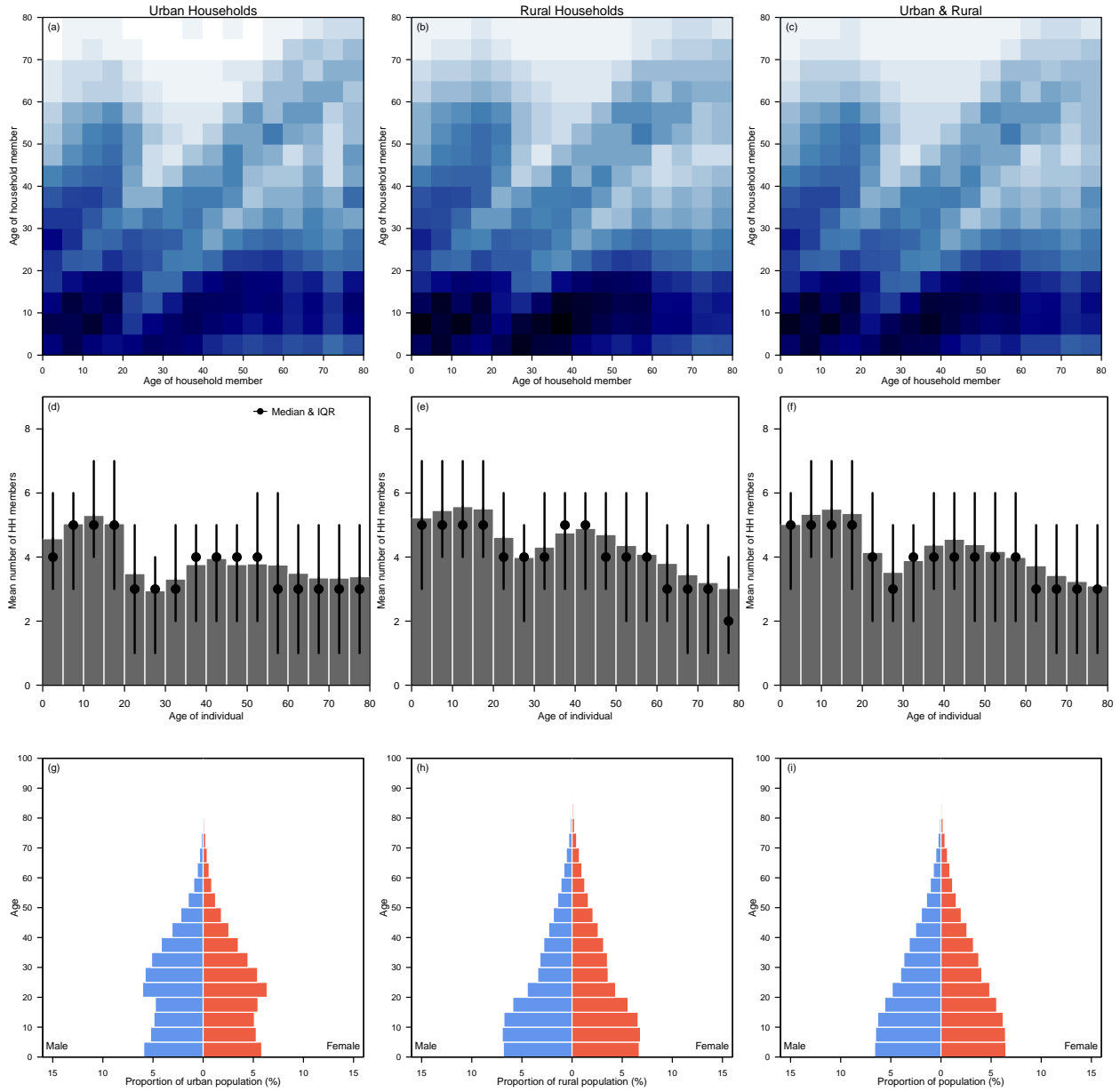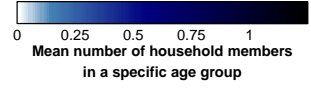
Empirical DHS household data: Kyrgyzstan
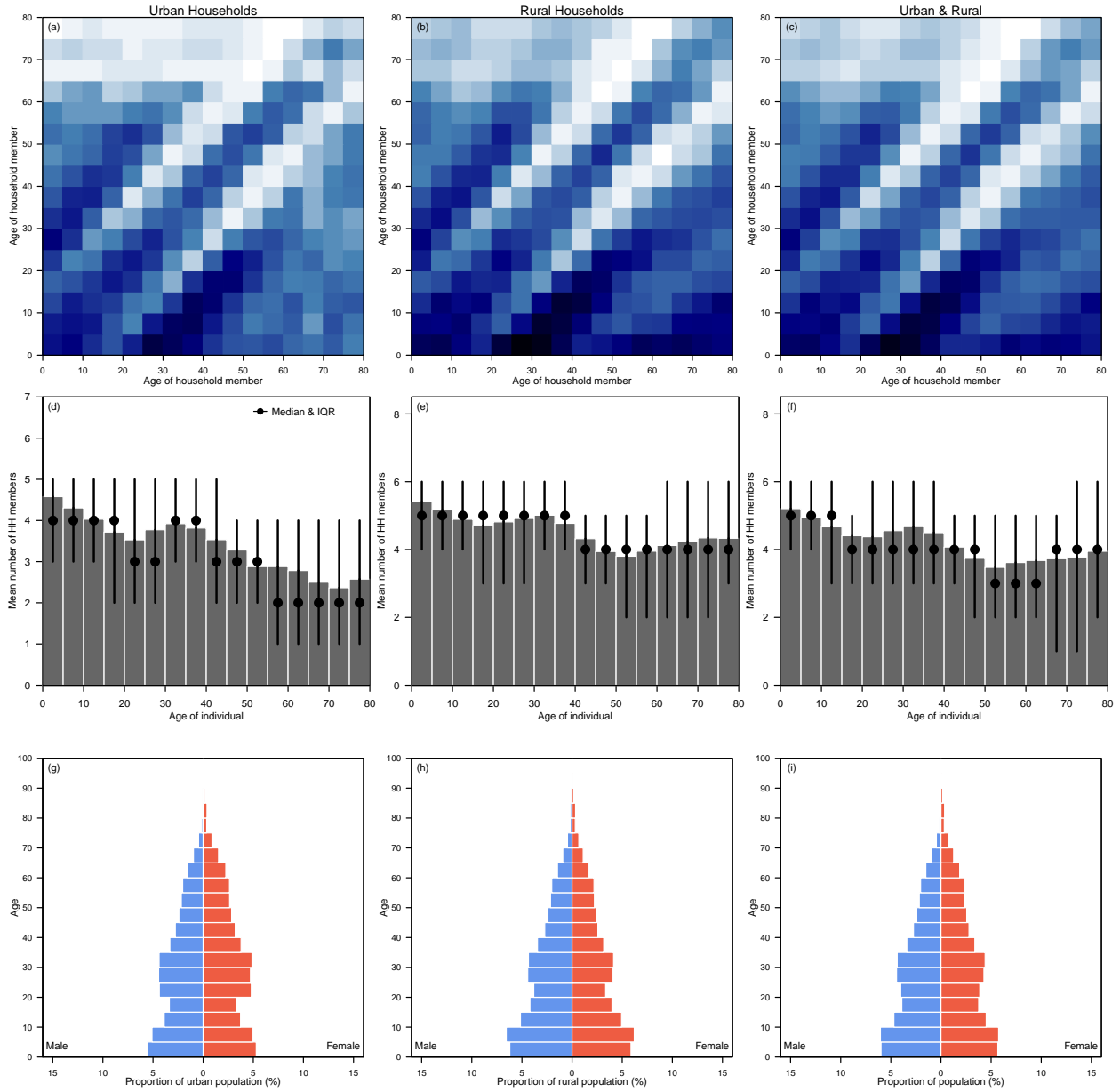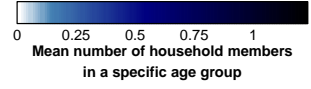
Empirical DHS household data: Lesotho
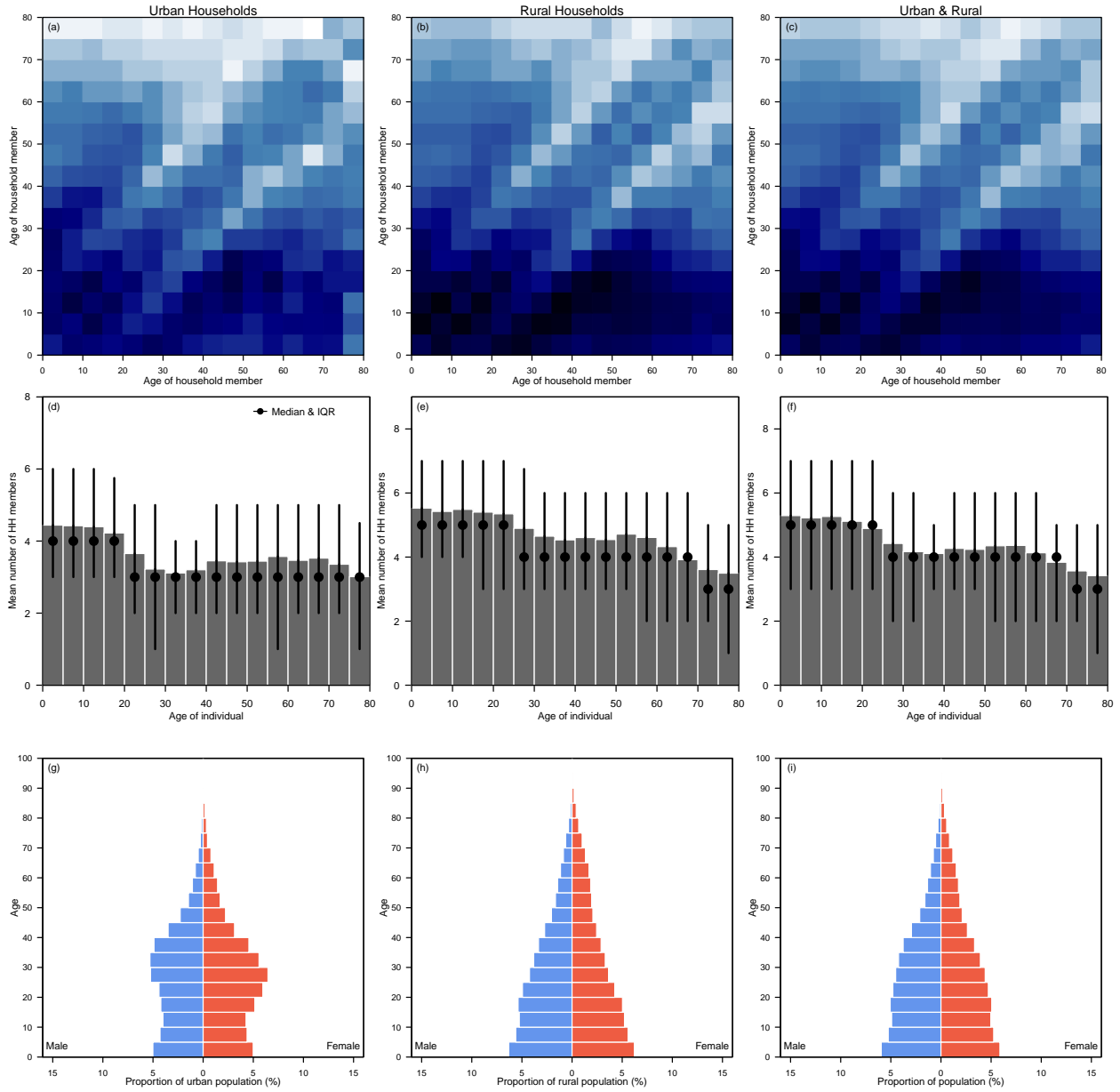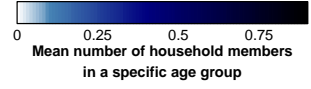
Empirical DHS household data: Liberia

Urban Households

Rural Households

Urban & Rural

Mean number of household members
in a specific age group
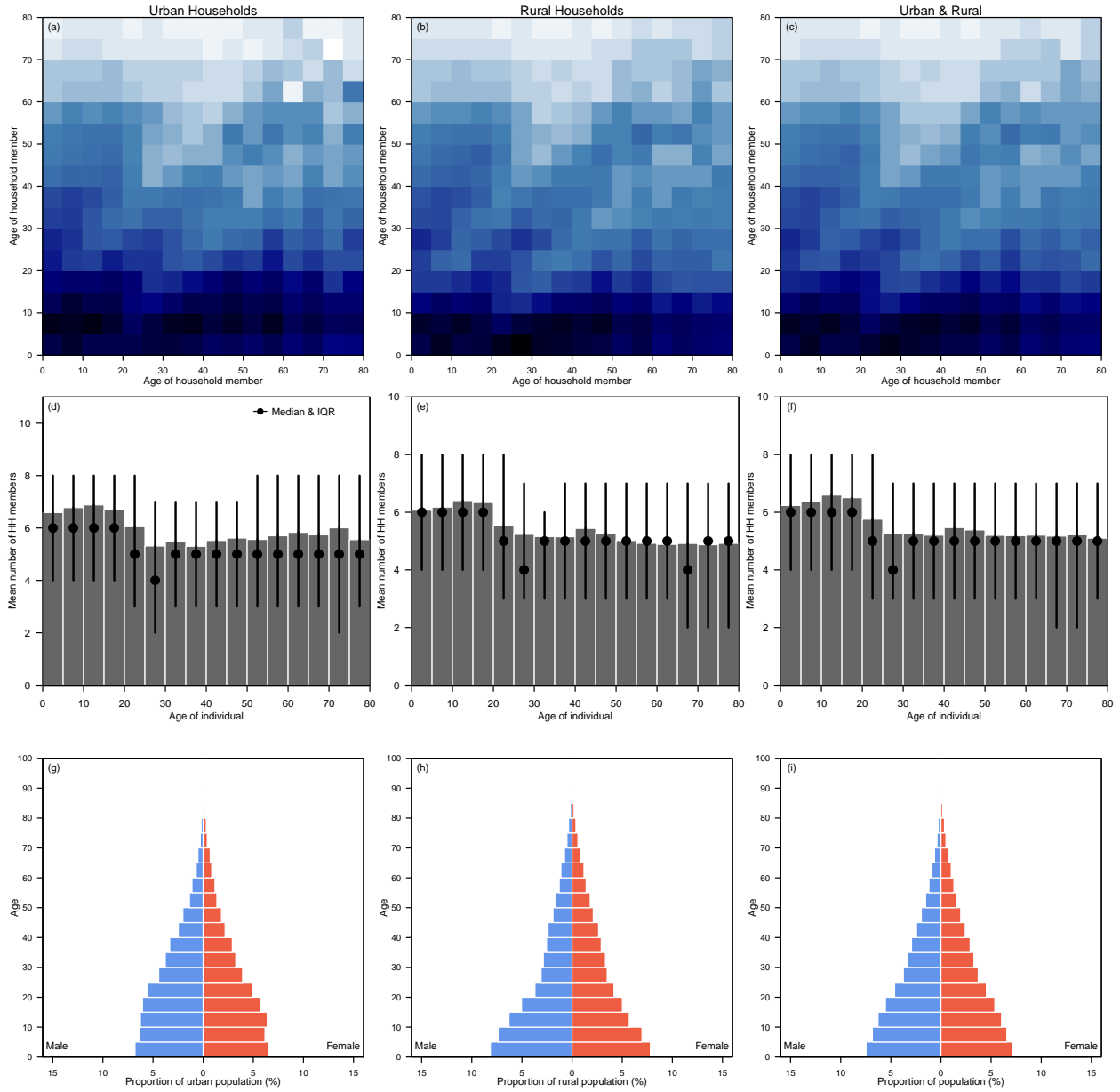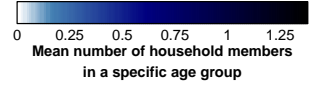
Empirical DHS household data: Malawi
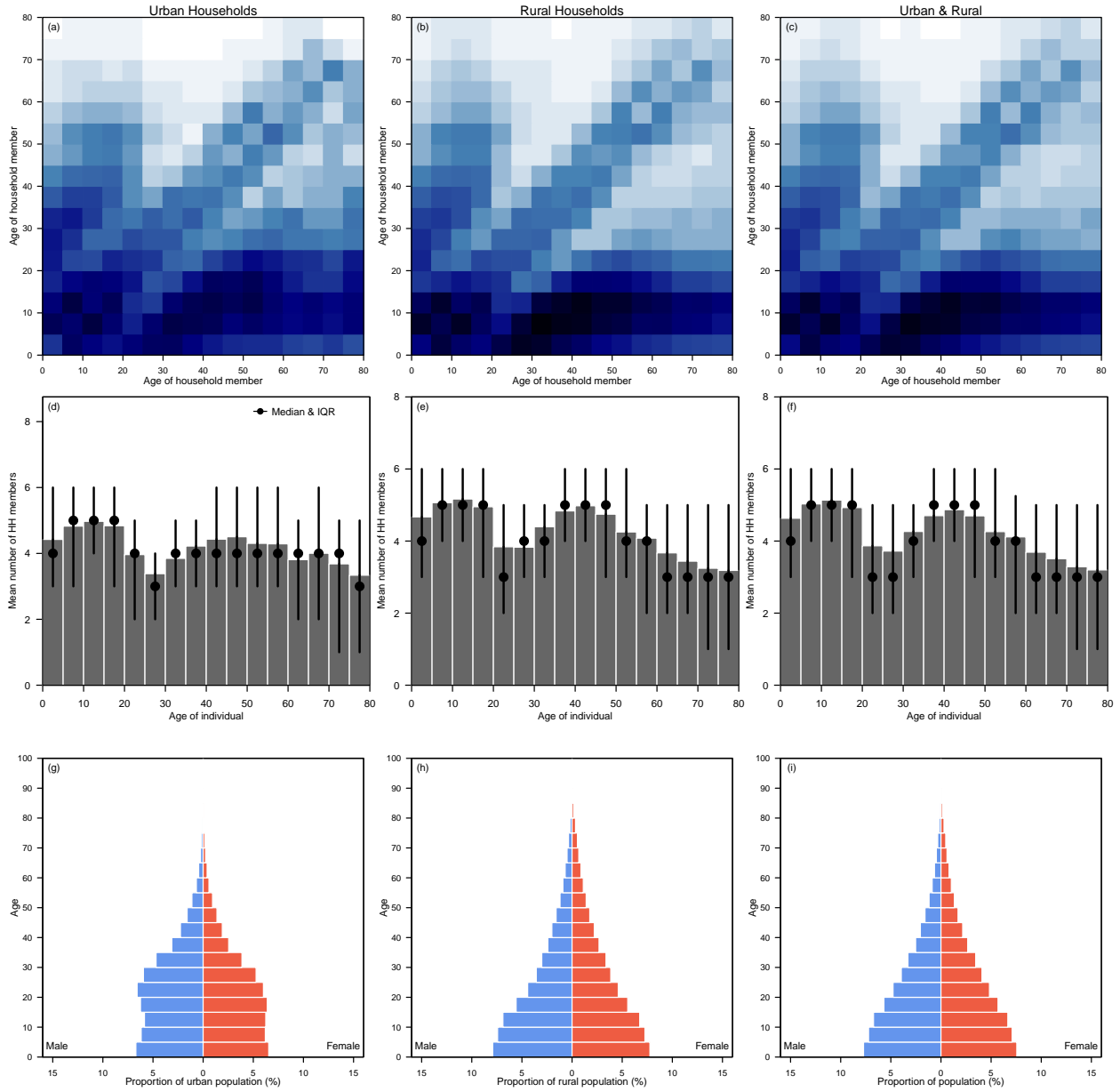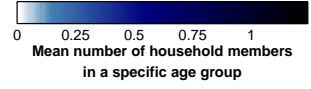
Empirical DHS household data: Maldives

Empirical DHS household data: Mali

Empirical DHS household data: Nepal

Mean number of household members in a specific age group

Urban Households

Rural Households

Urban & Rural
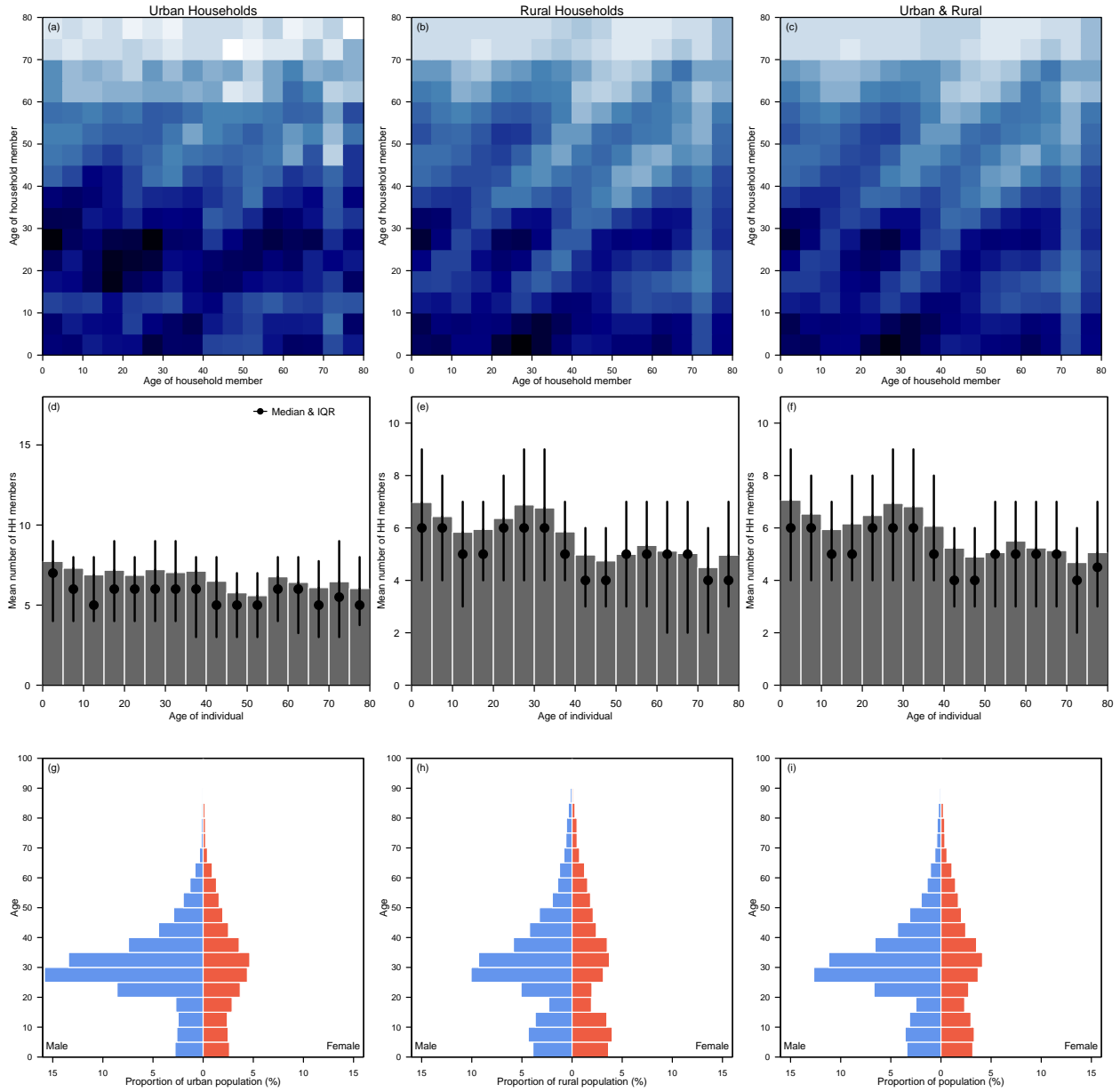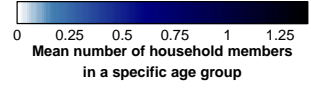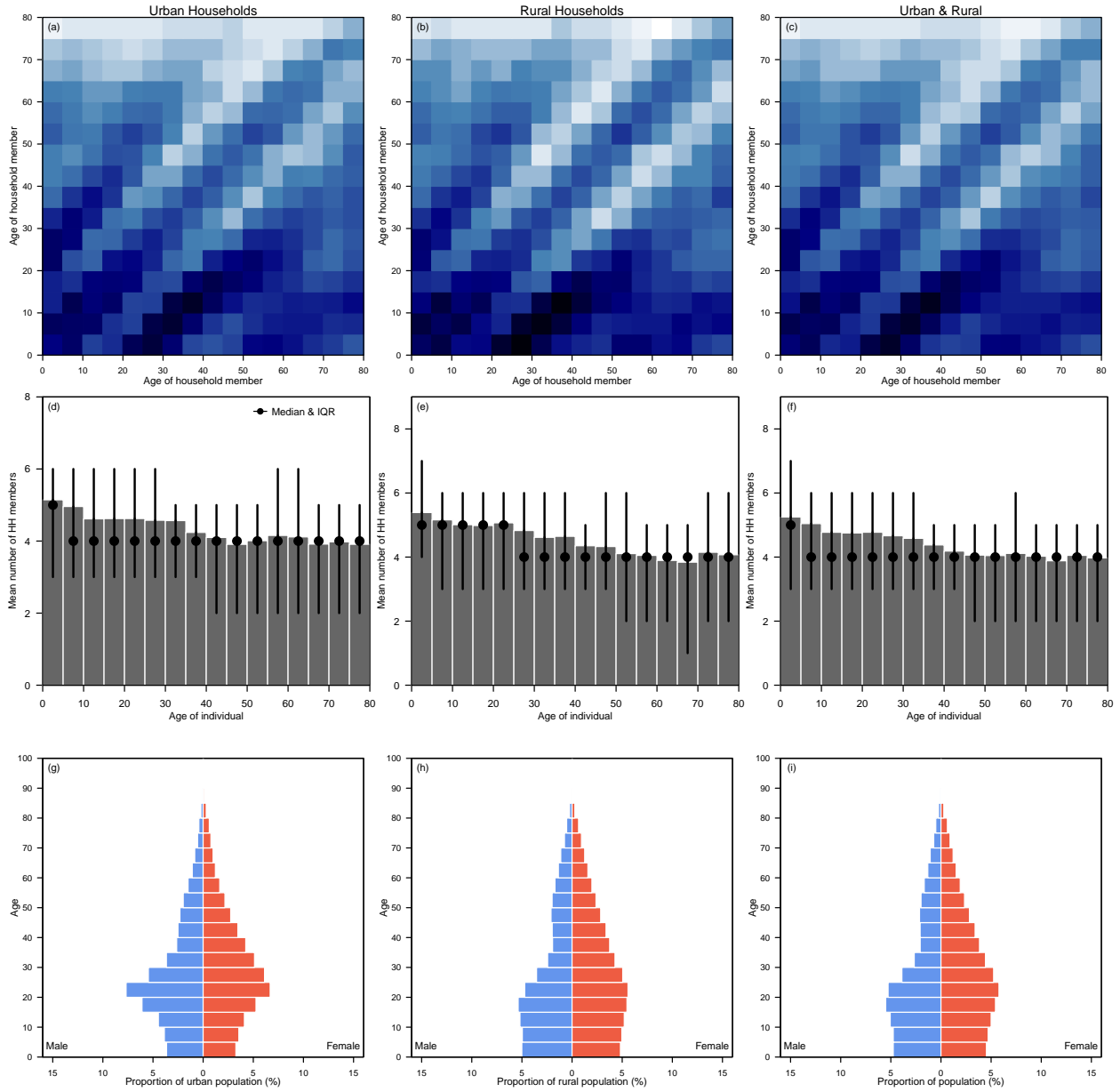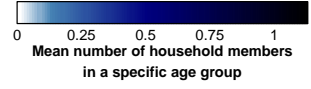
Empirical DHS household data: Niger

Empirical DHS household data: Nigeria

Mean number of household members in a specific age group
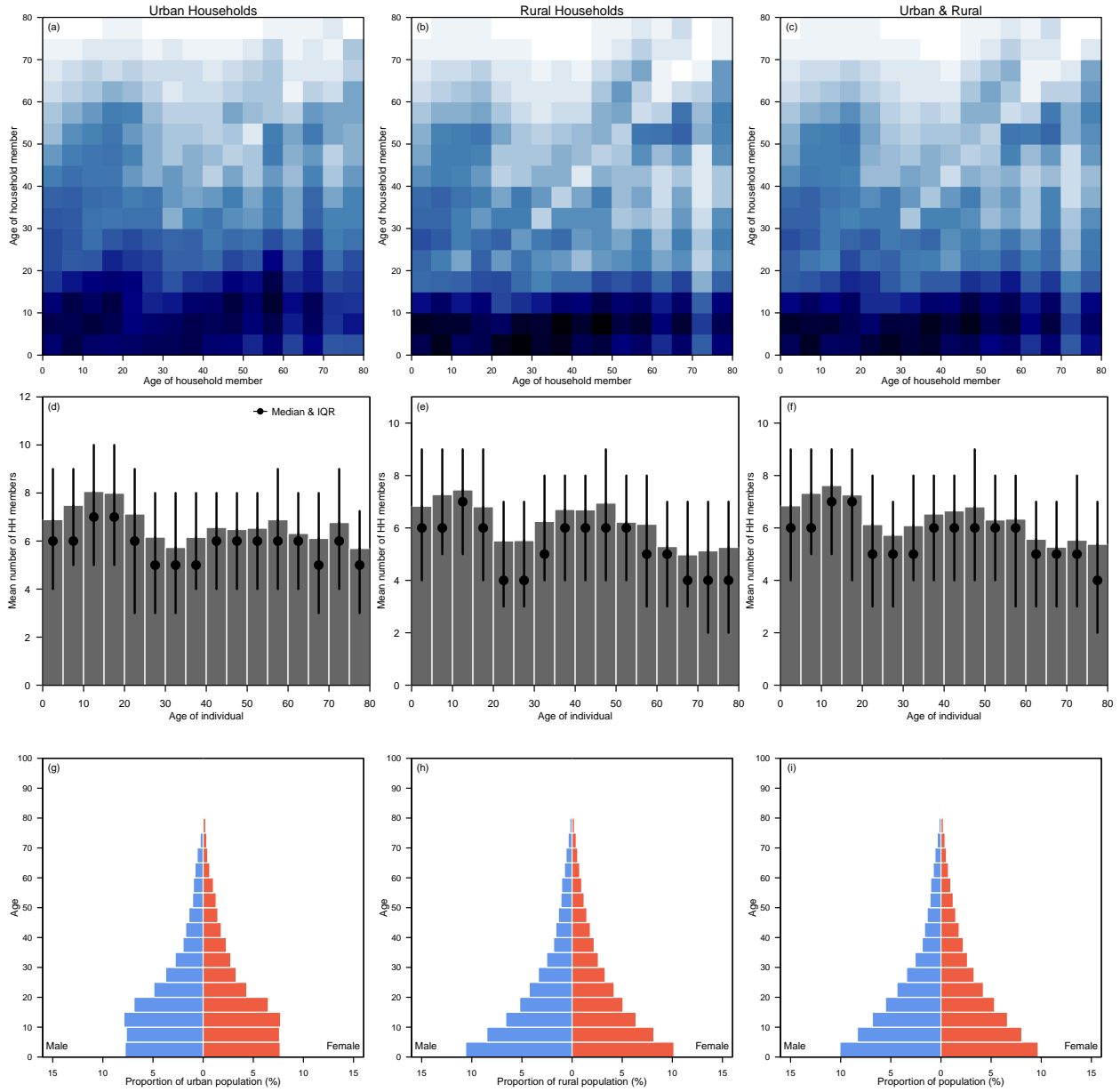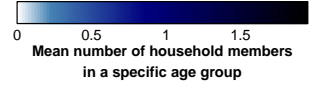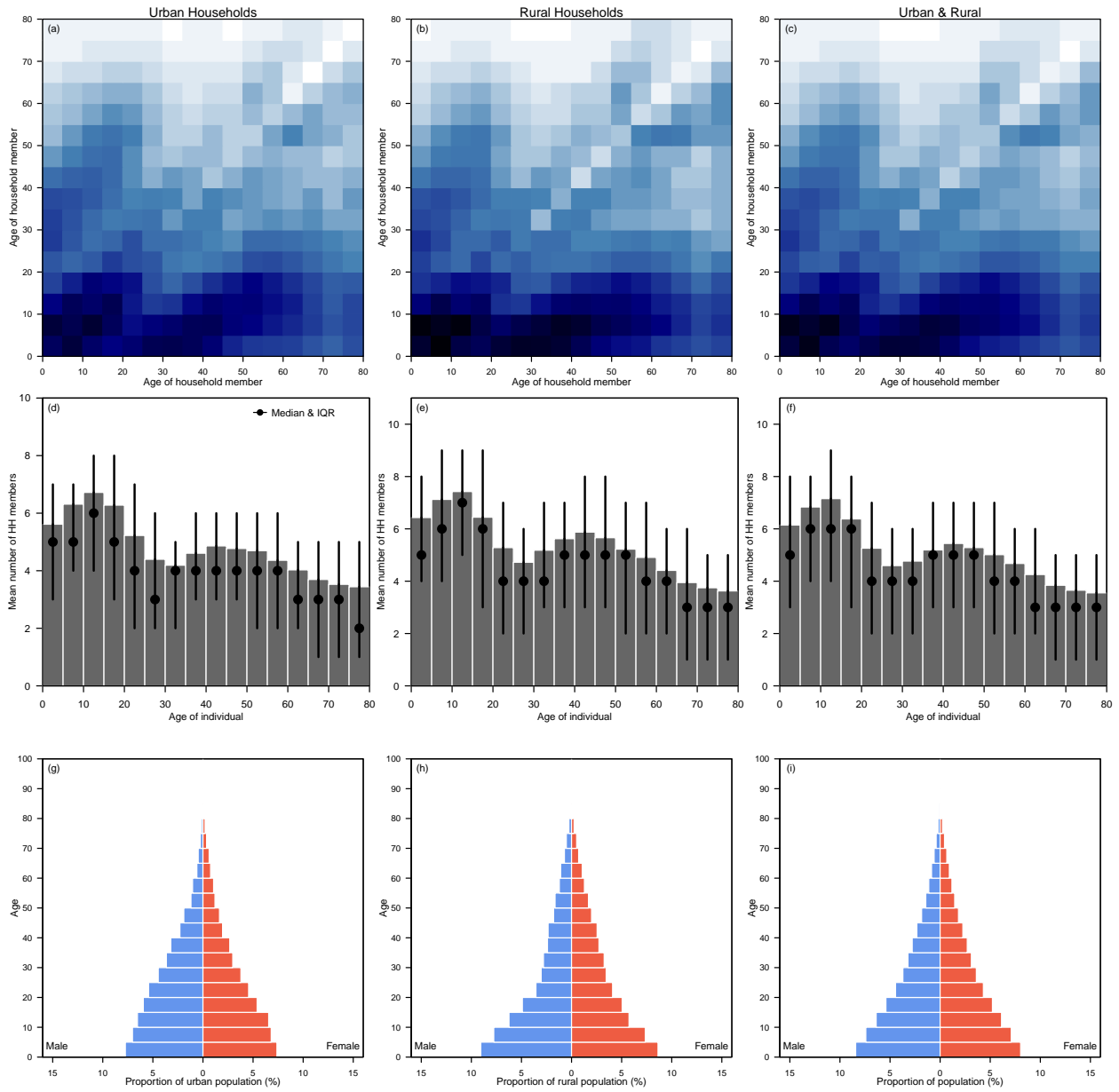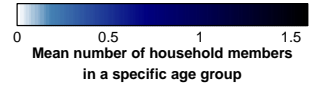
Empirical DHS household data: Pakistan

Mean number of household members in a specific age group
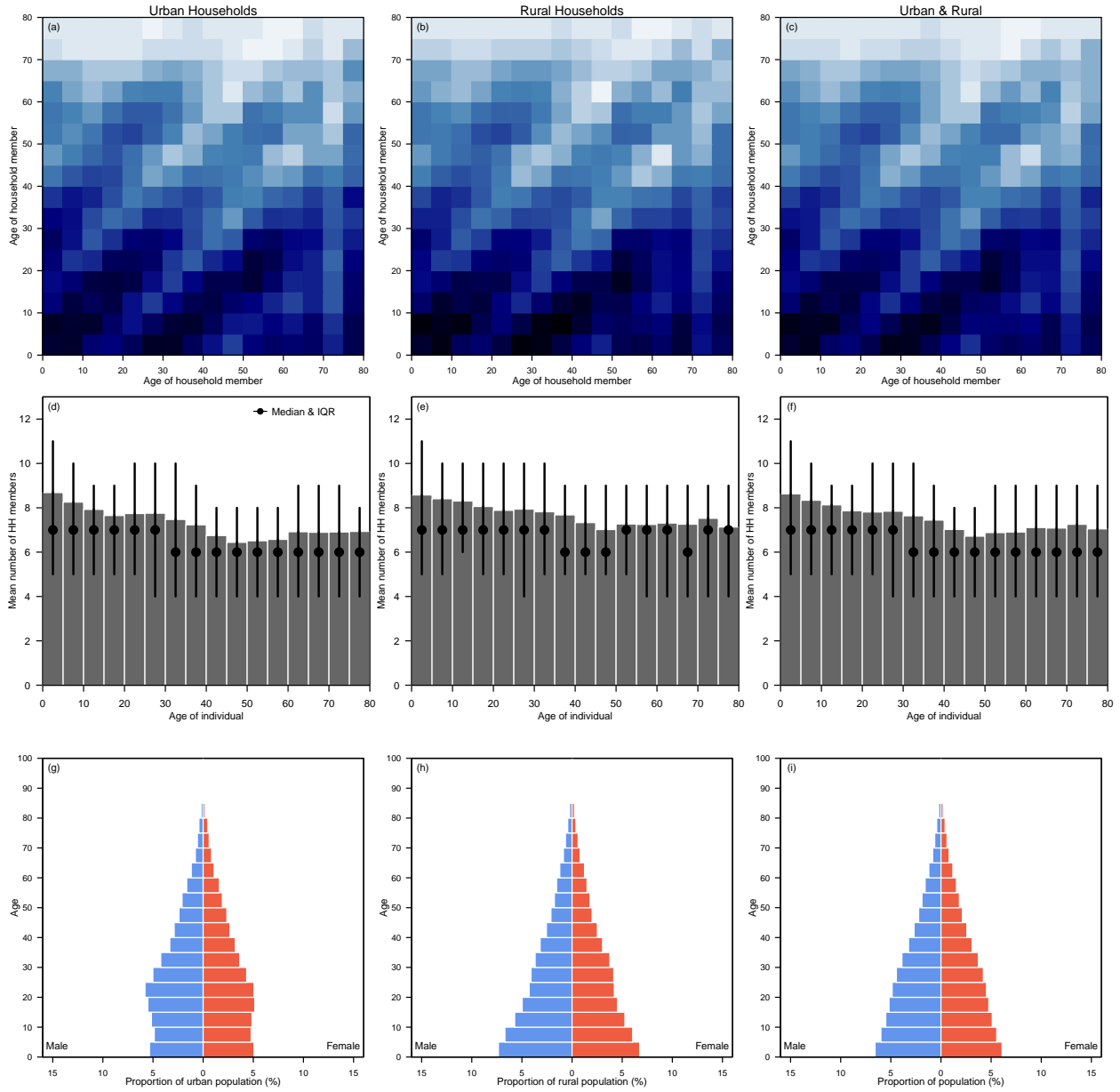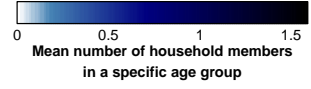
Empirical DHS household data: Peru
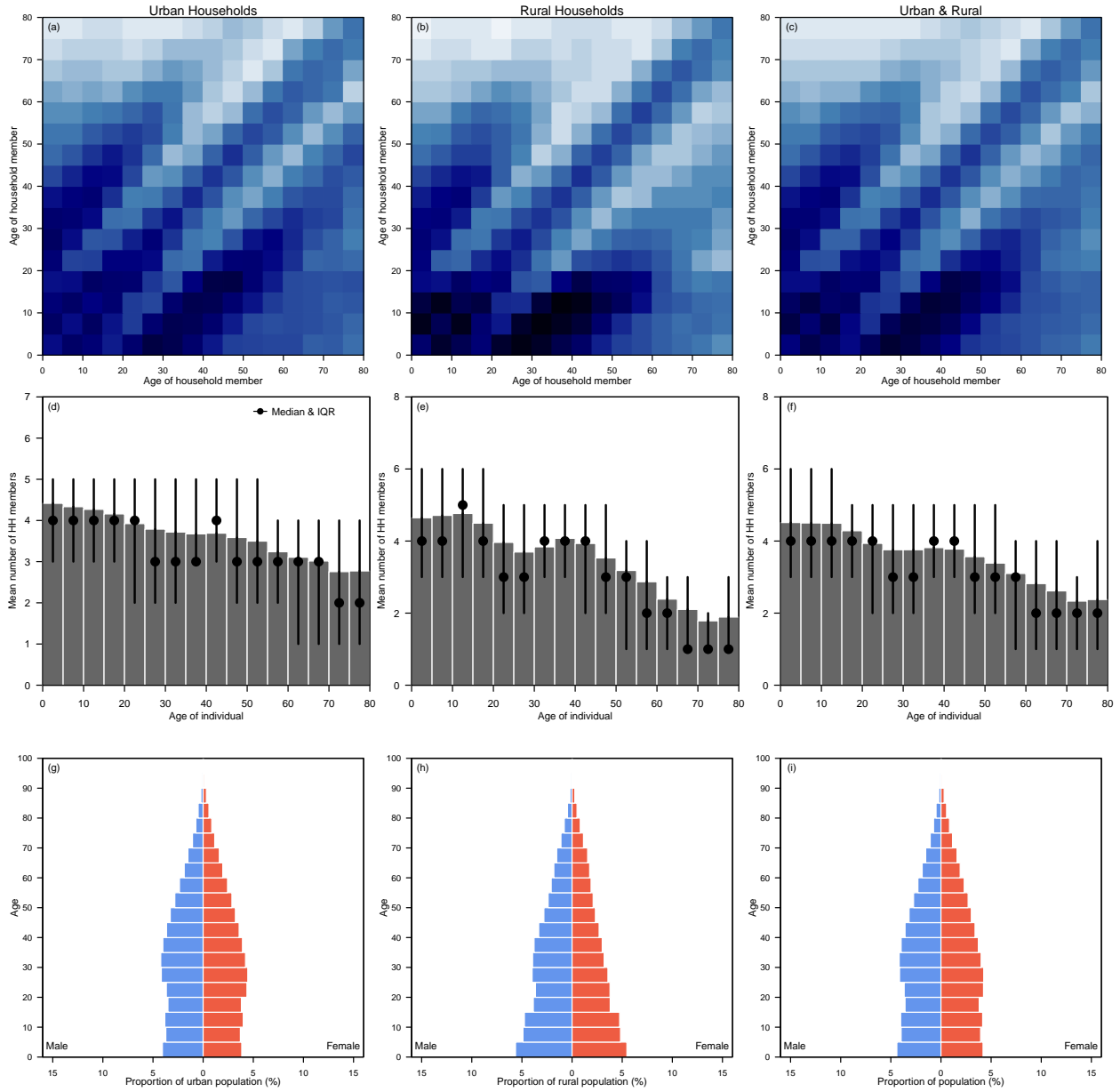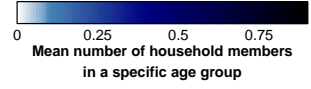
Empirical DHS household data: Philippines
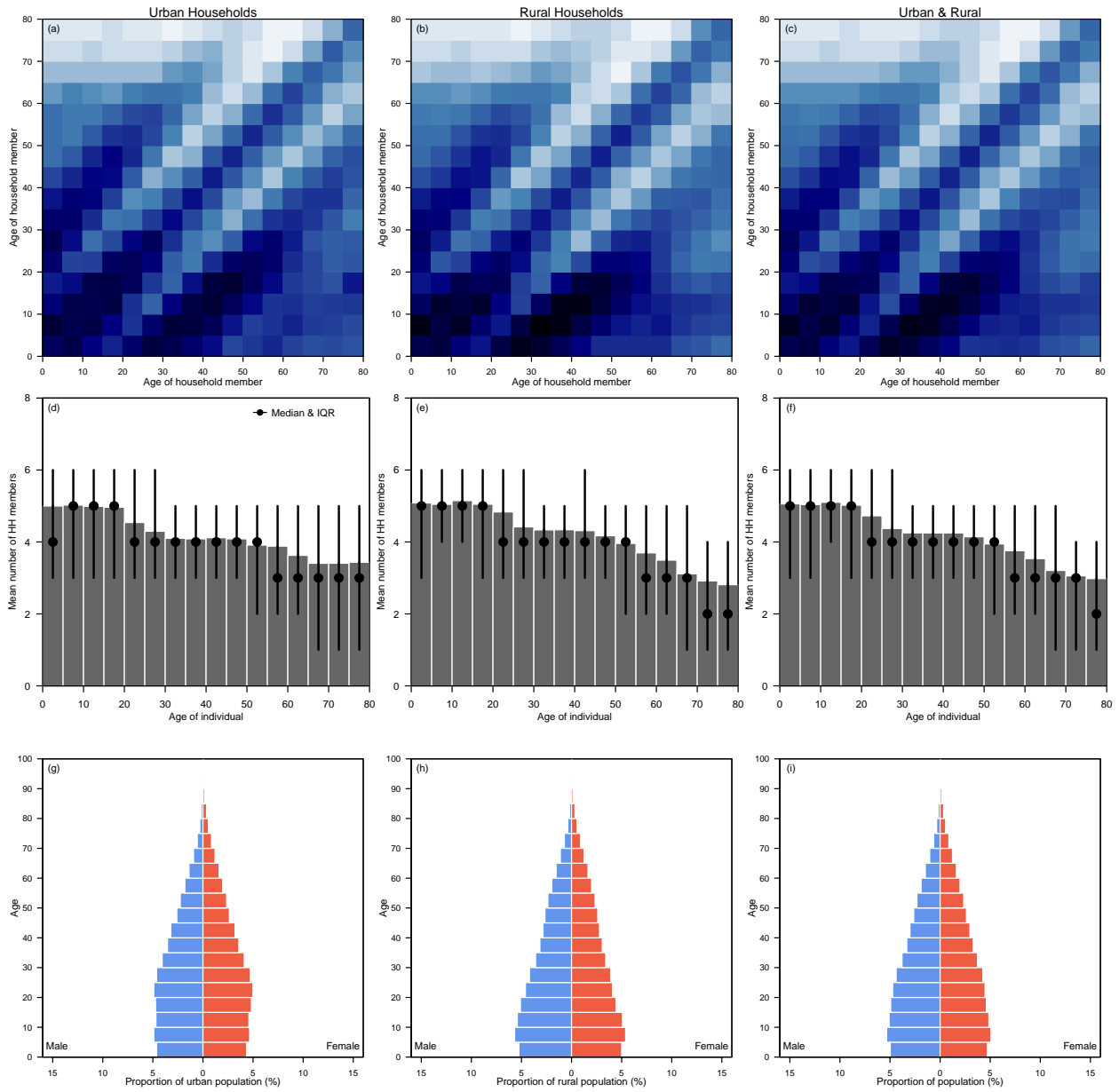
Empirical DHS household data: Senegal
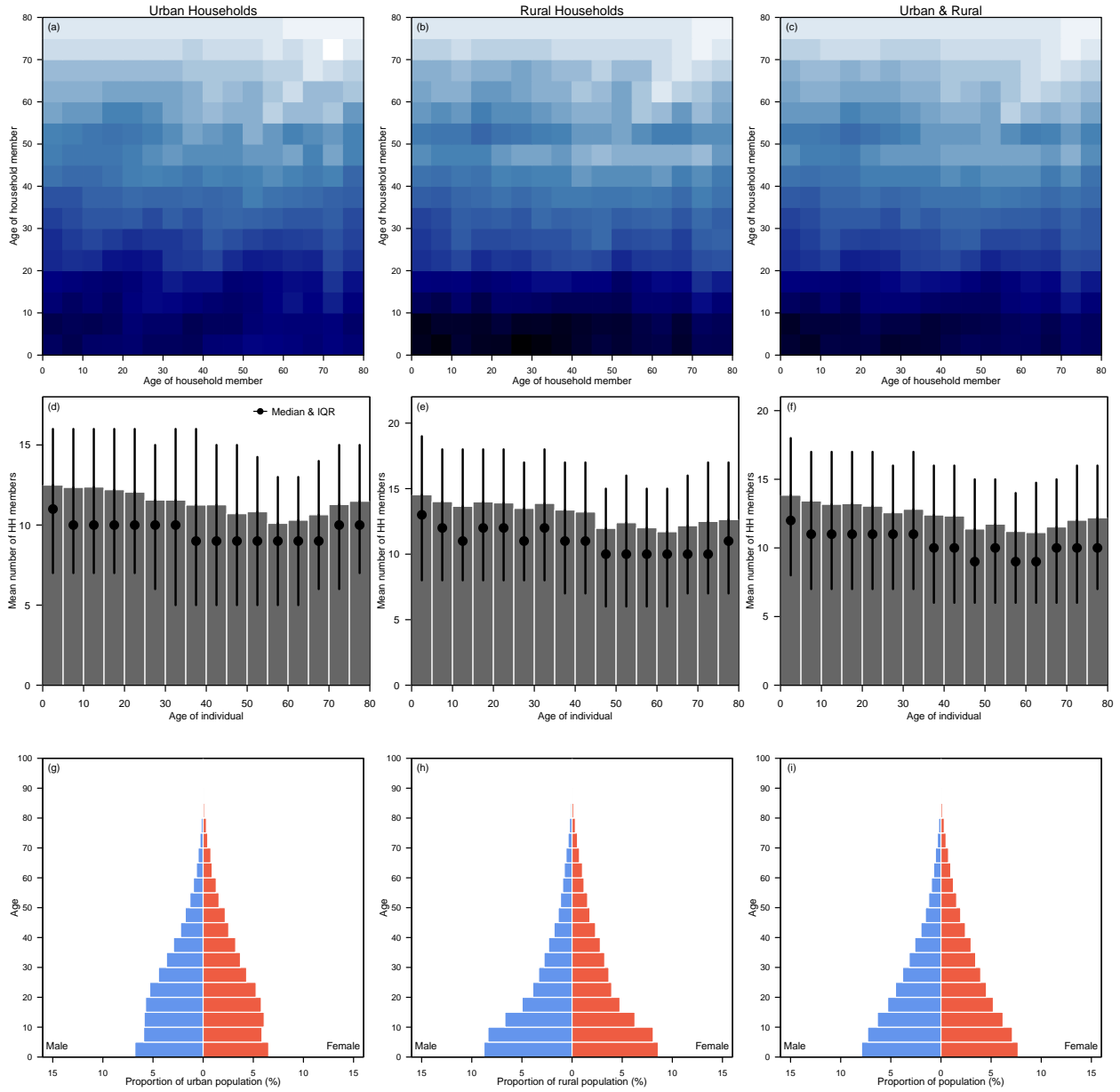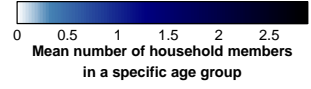
Empirical DHS household data: Sierra Leone

Empirical DHS household data: South Africa
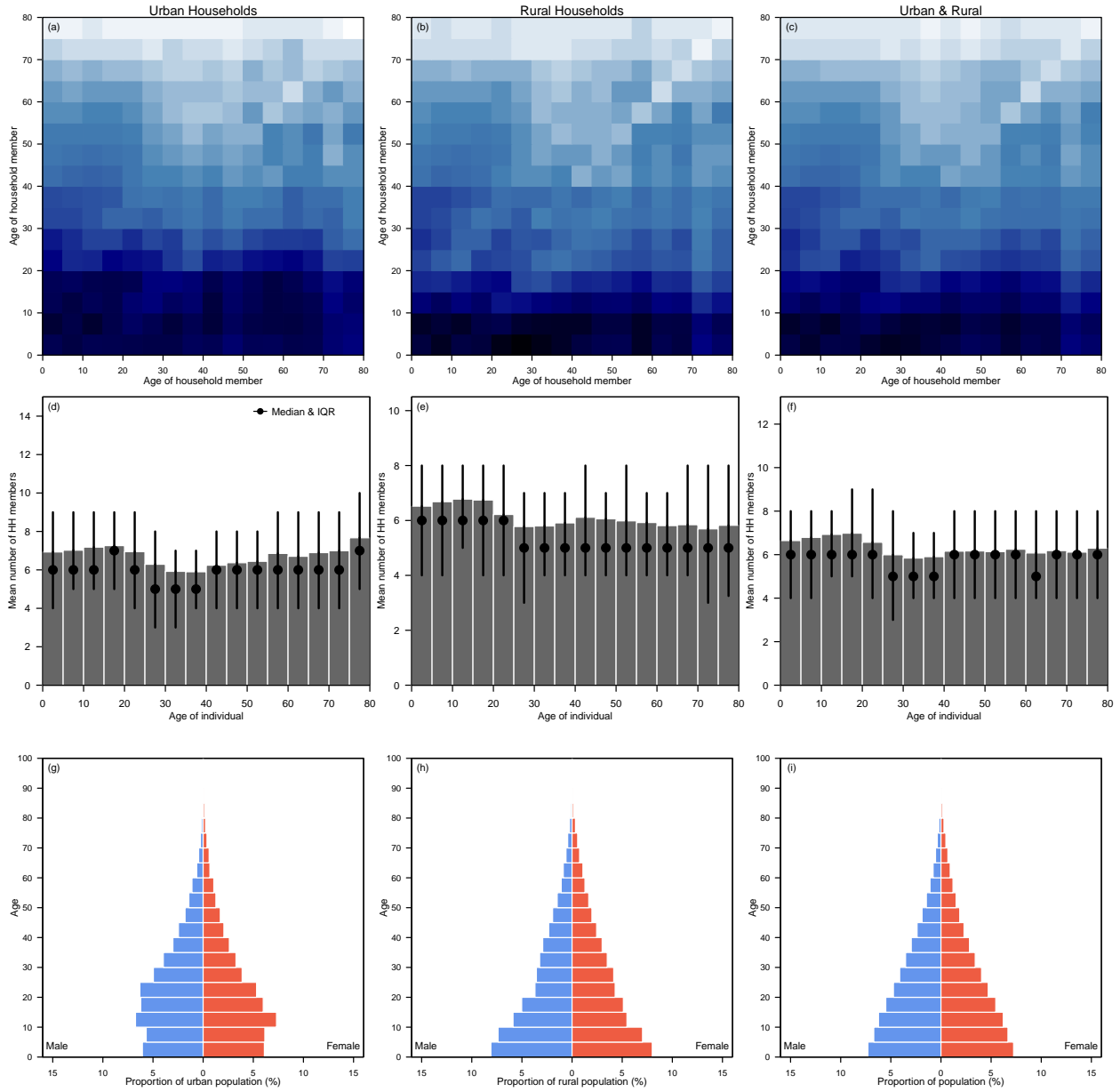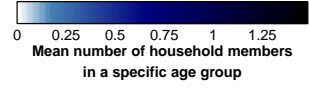
Empirical DHS household data: Timor–Leste
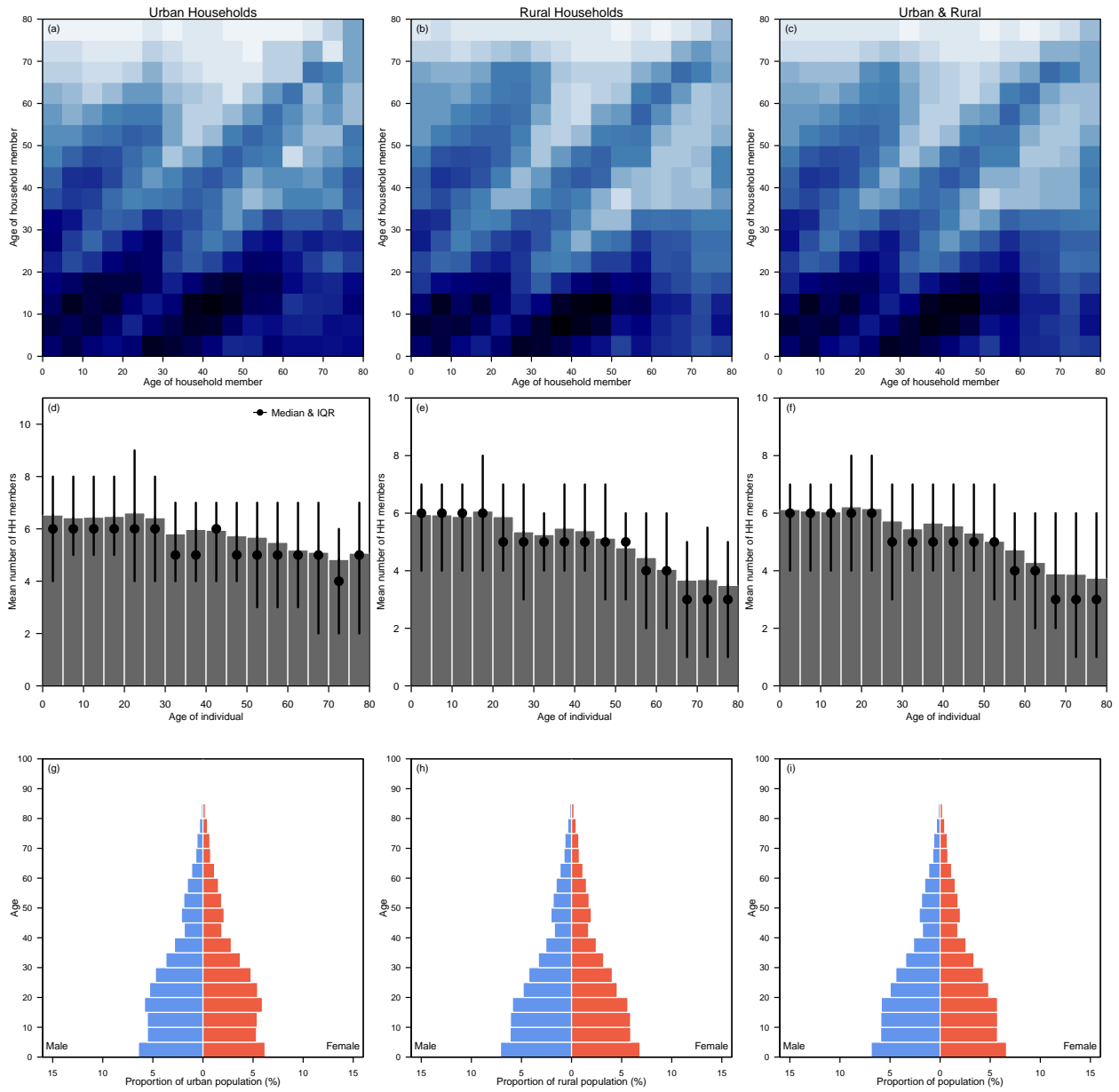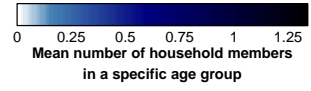
Empirical DHS household data: Togo
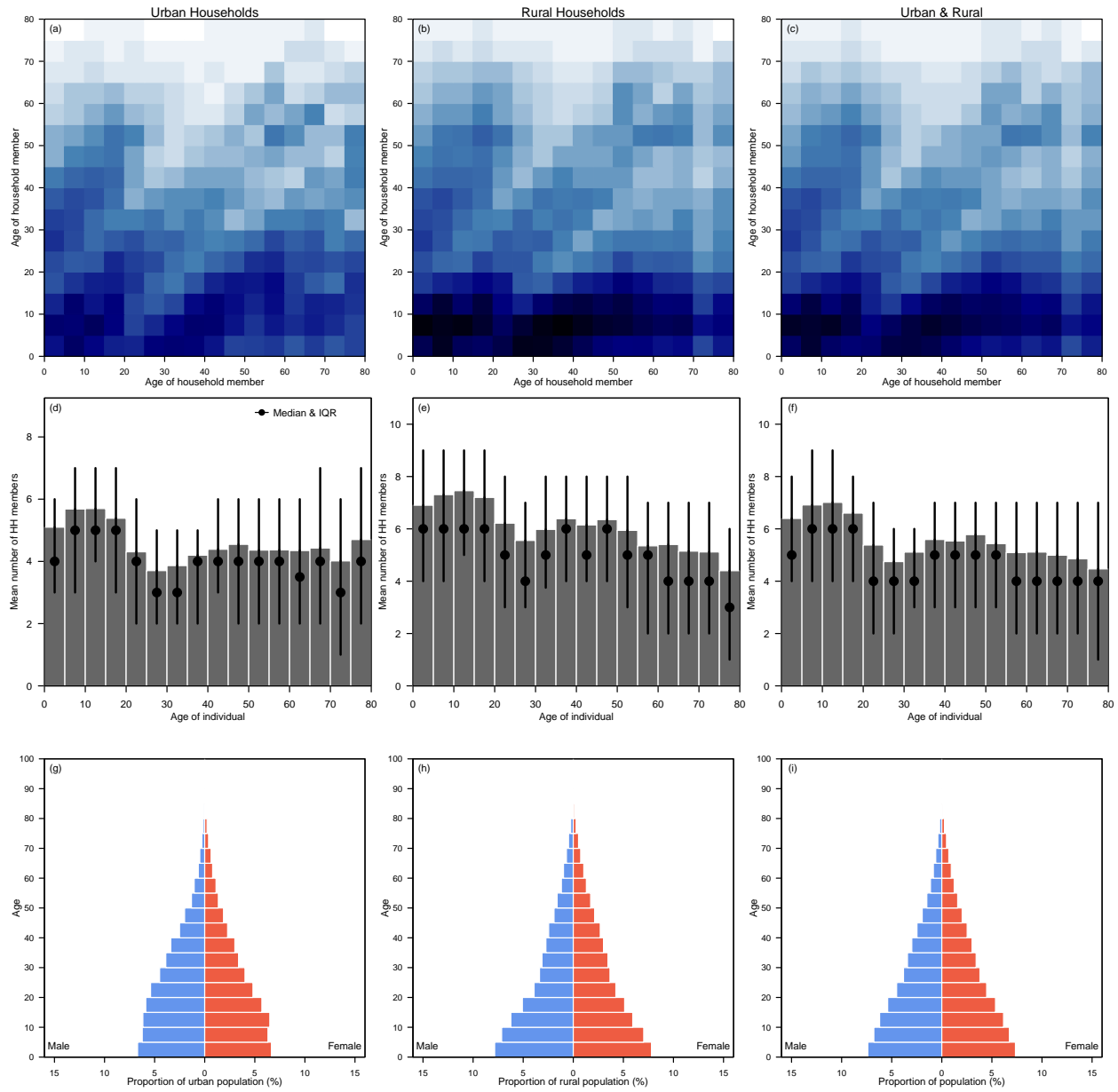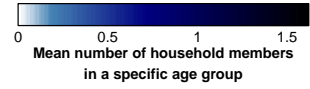
Empirical DHS household data: Uganda
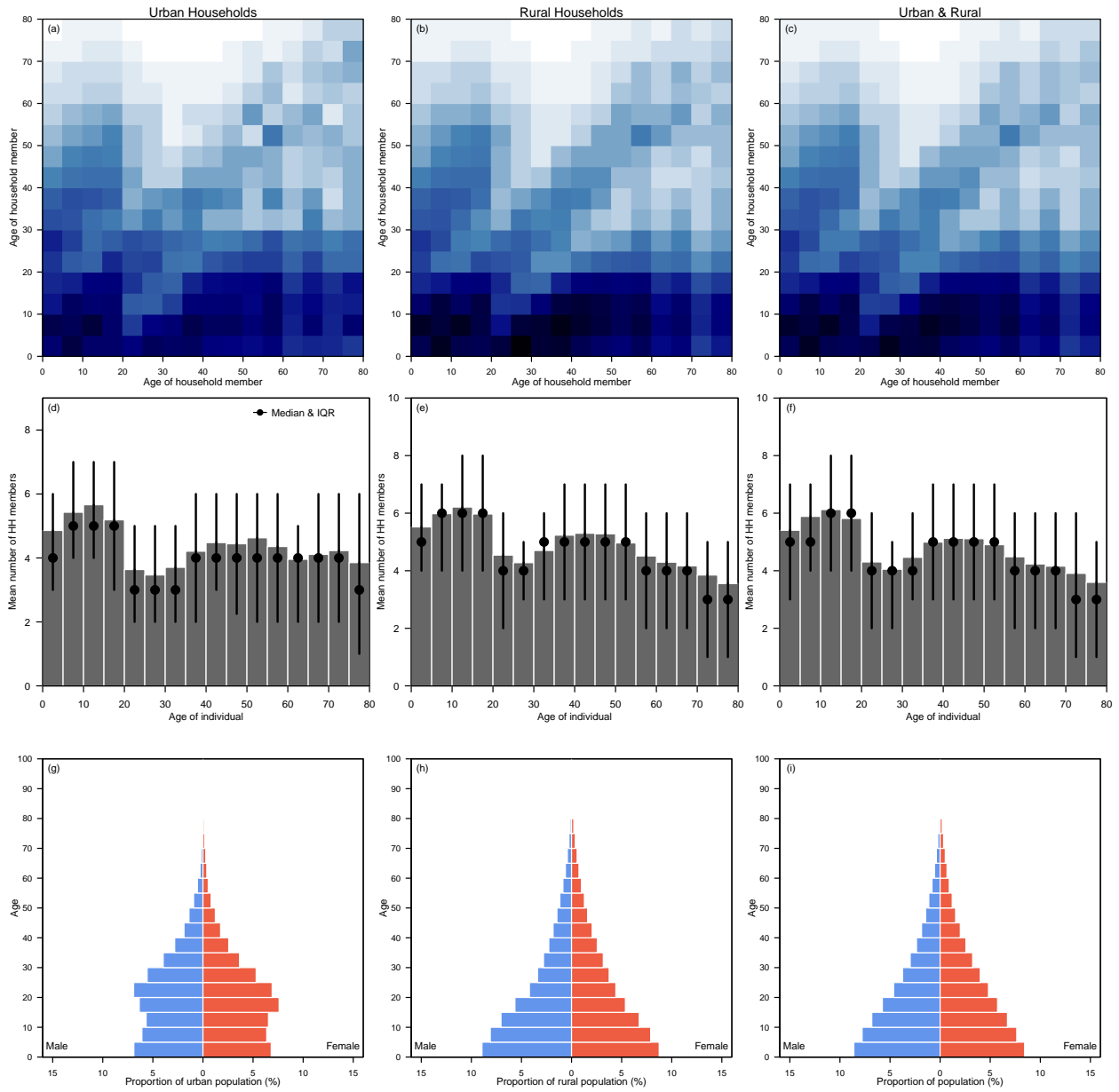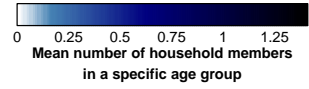
Empirical DHS household data: Viet Nam
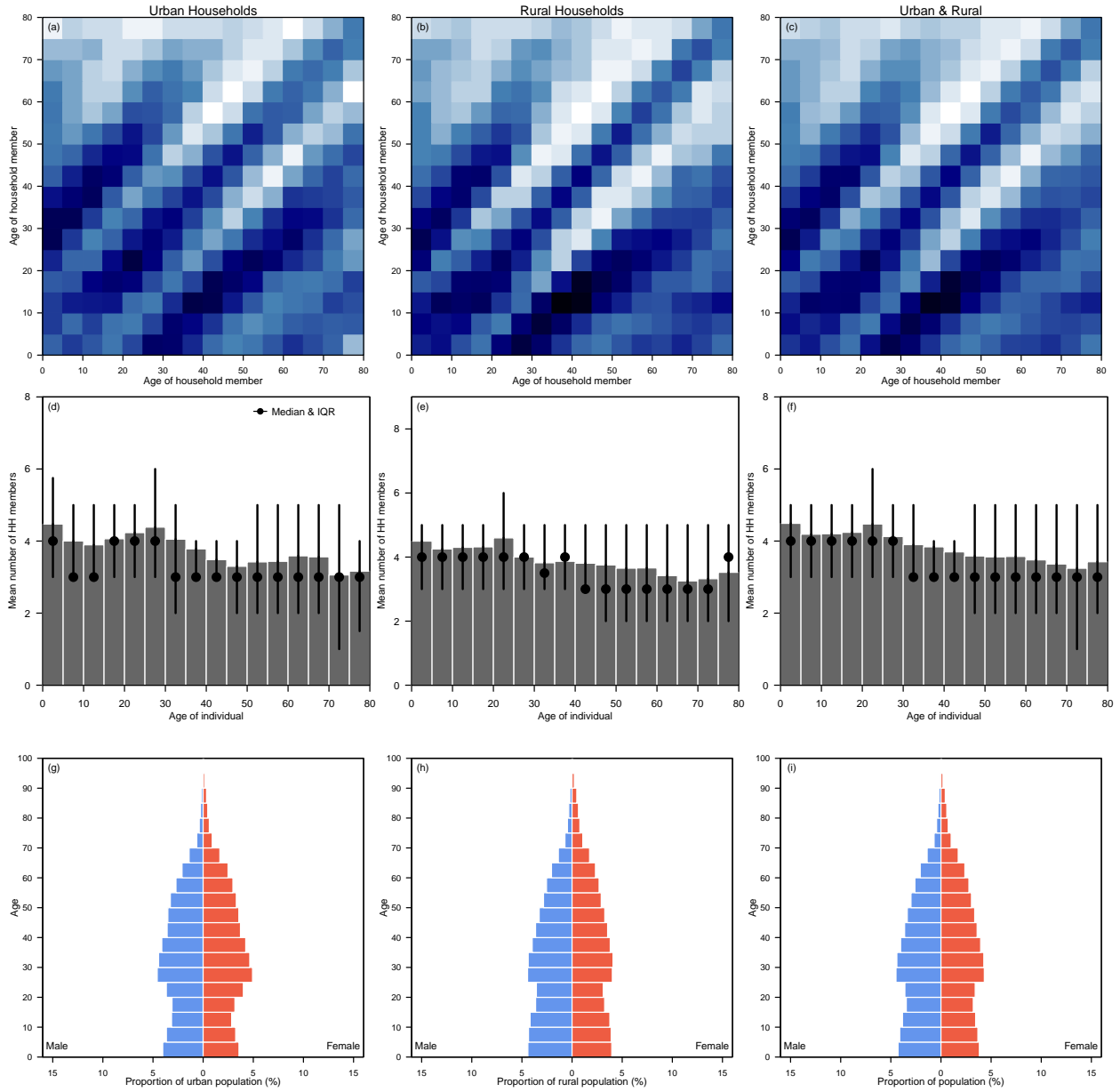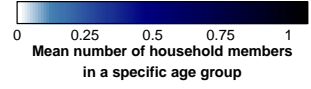
Empirical DHS household data: Zambia
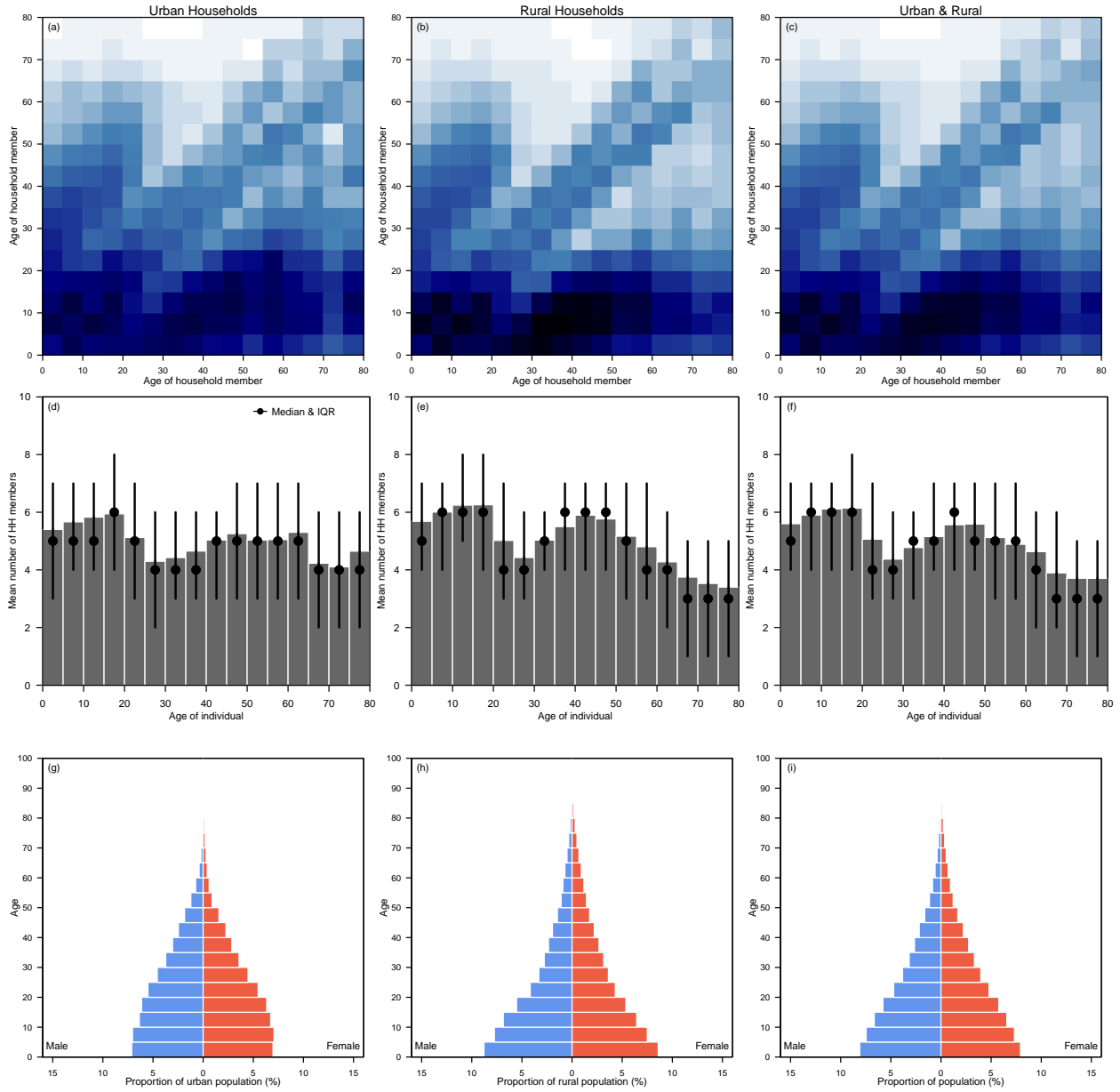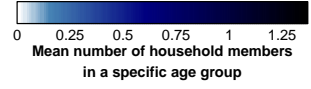
Empirical DHS household data: Zimbabwe
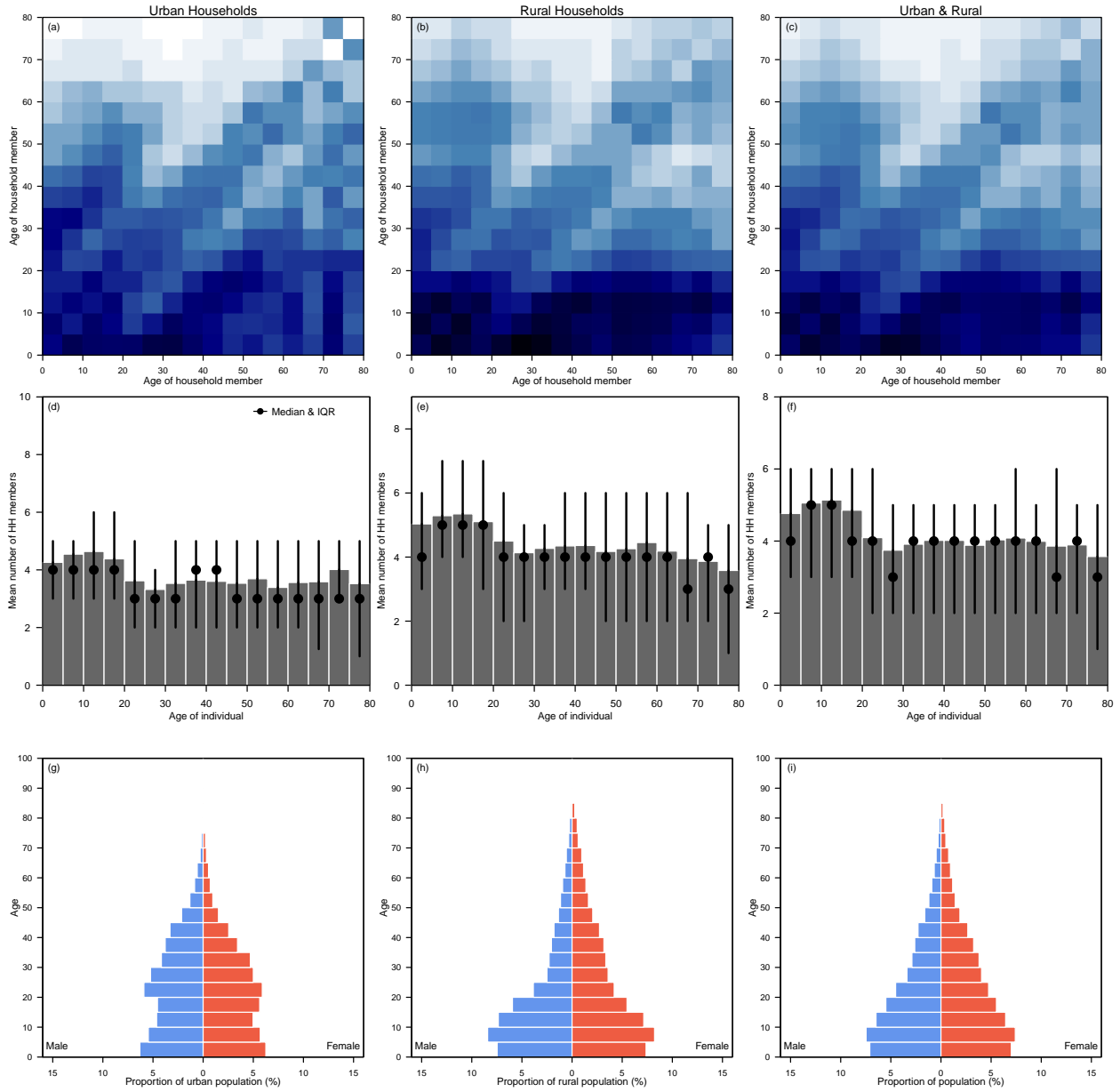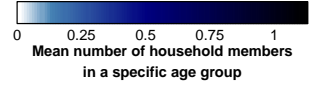
2.3 Updated age- and location-specifc contact matrices

The 2017 synthetic matrices provide validated approximations to age-and-location-specific contact matrices for 152 geographical regions. We have therefore updated these matrices with the most recent data—Demographic Household Surveys, World Bank and UN Population Division—extending the coverage to 177 geographical locations, covering 97.2% of the world's population.

Afghanistan
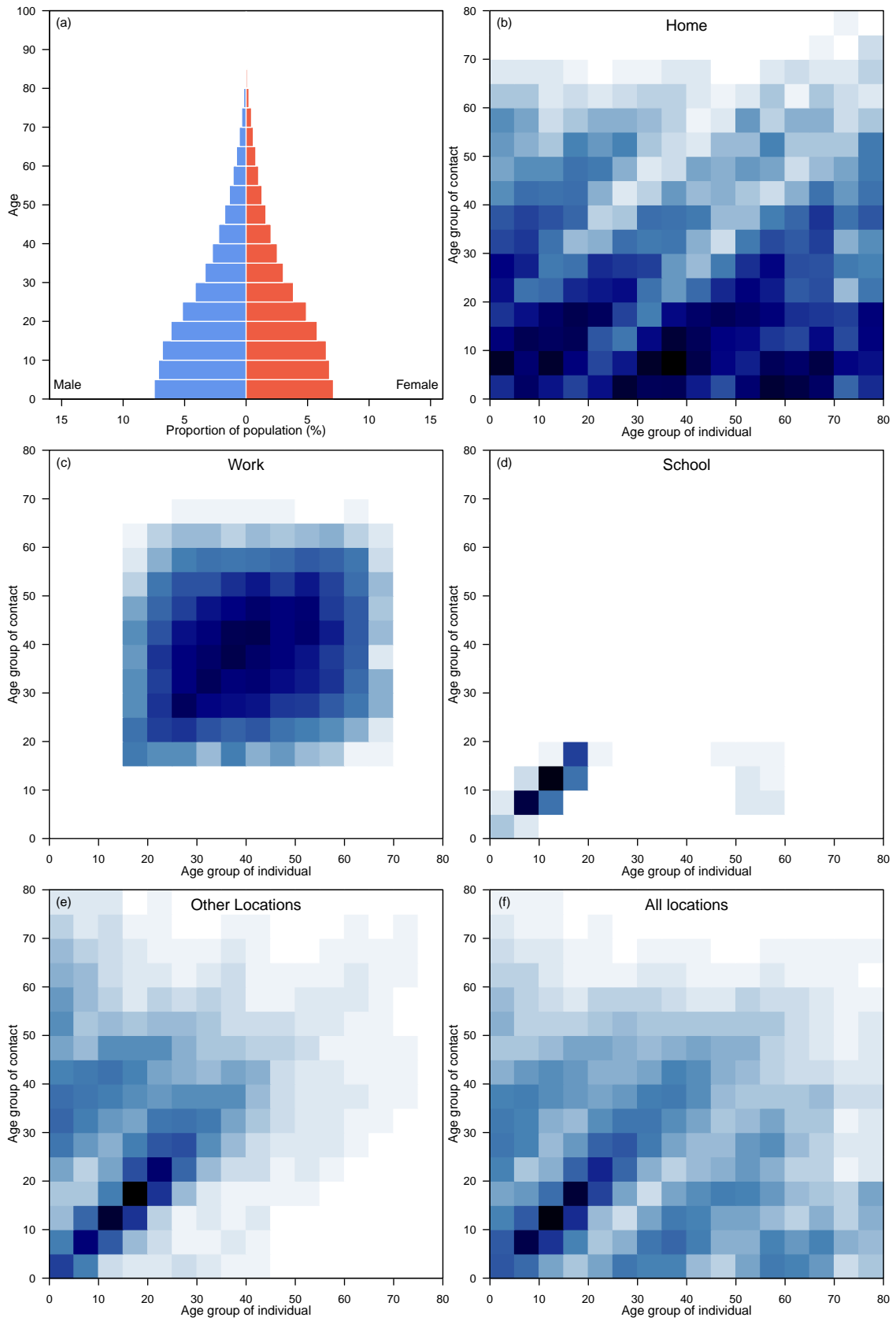
Albania
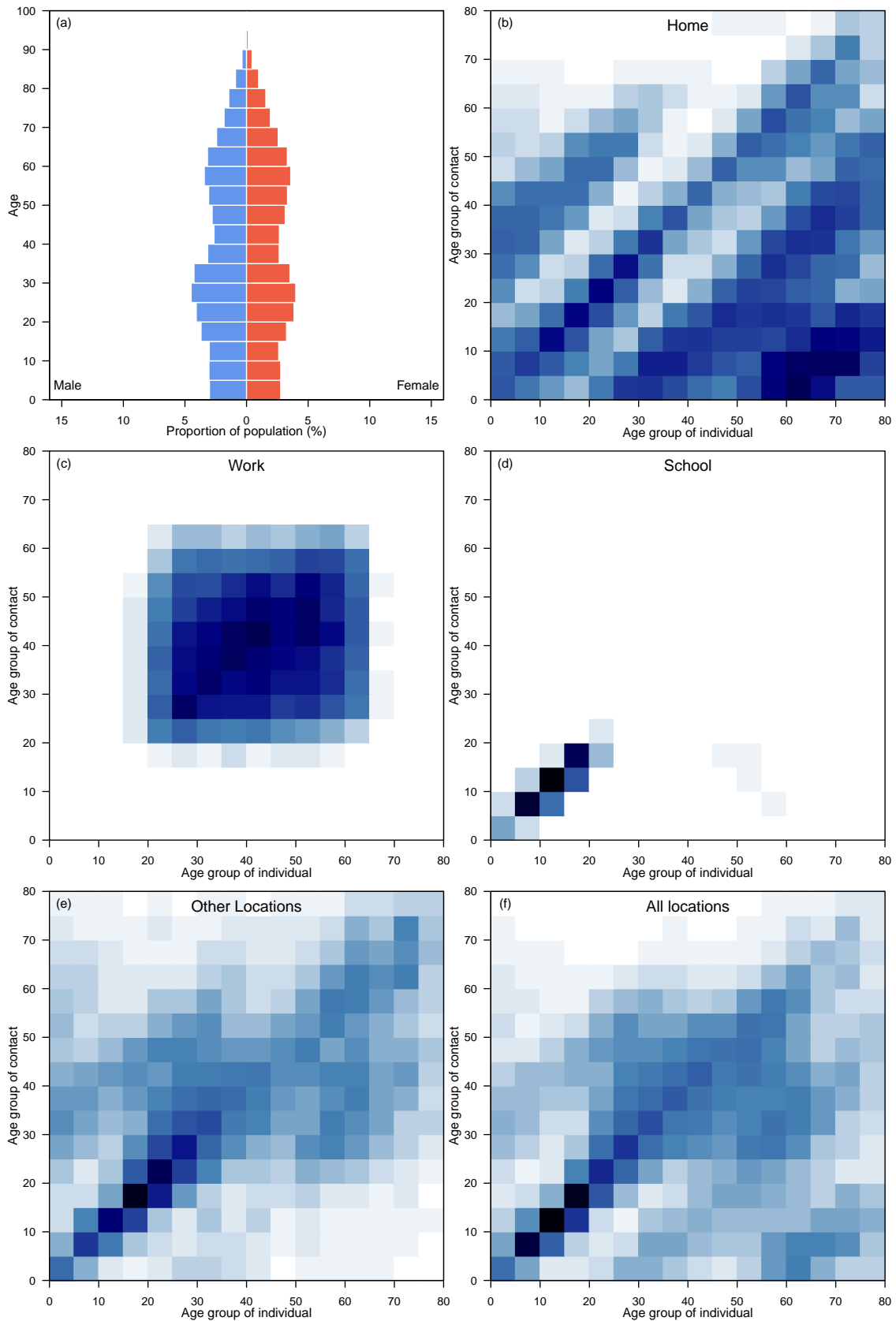
Algeria
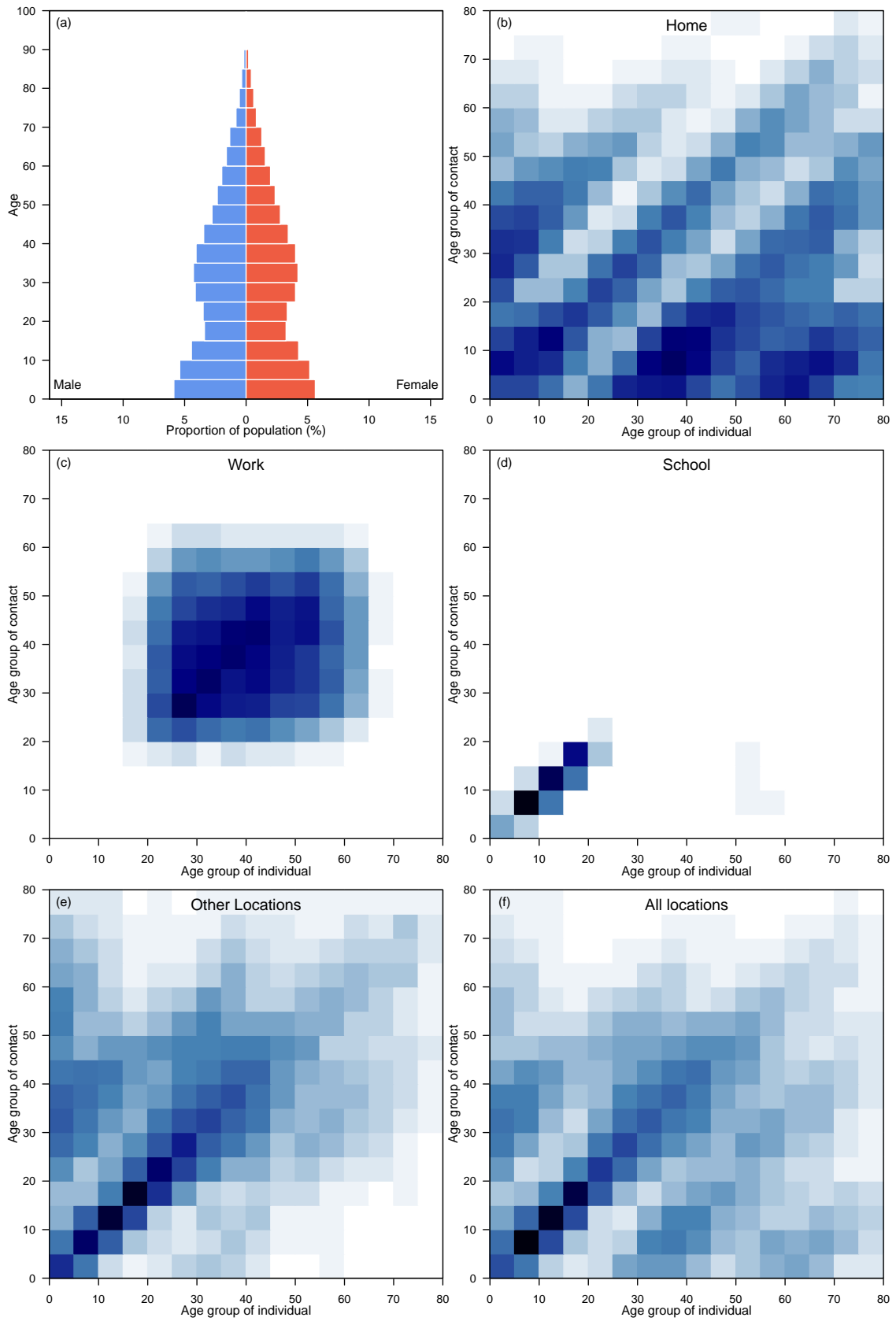
Angola
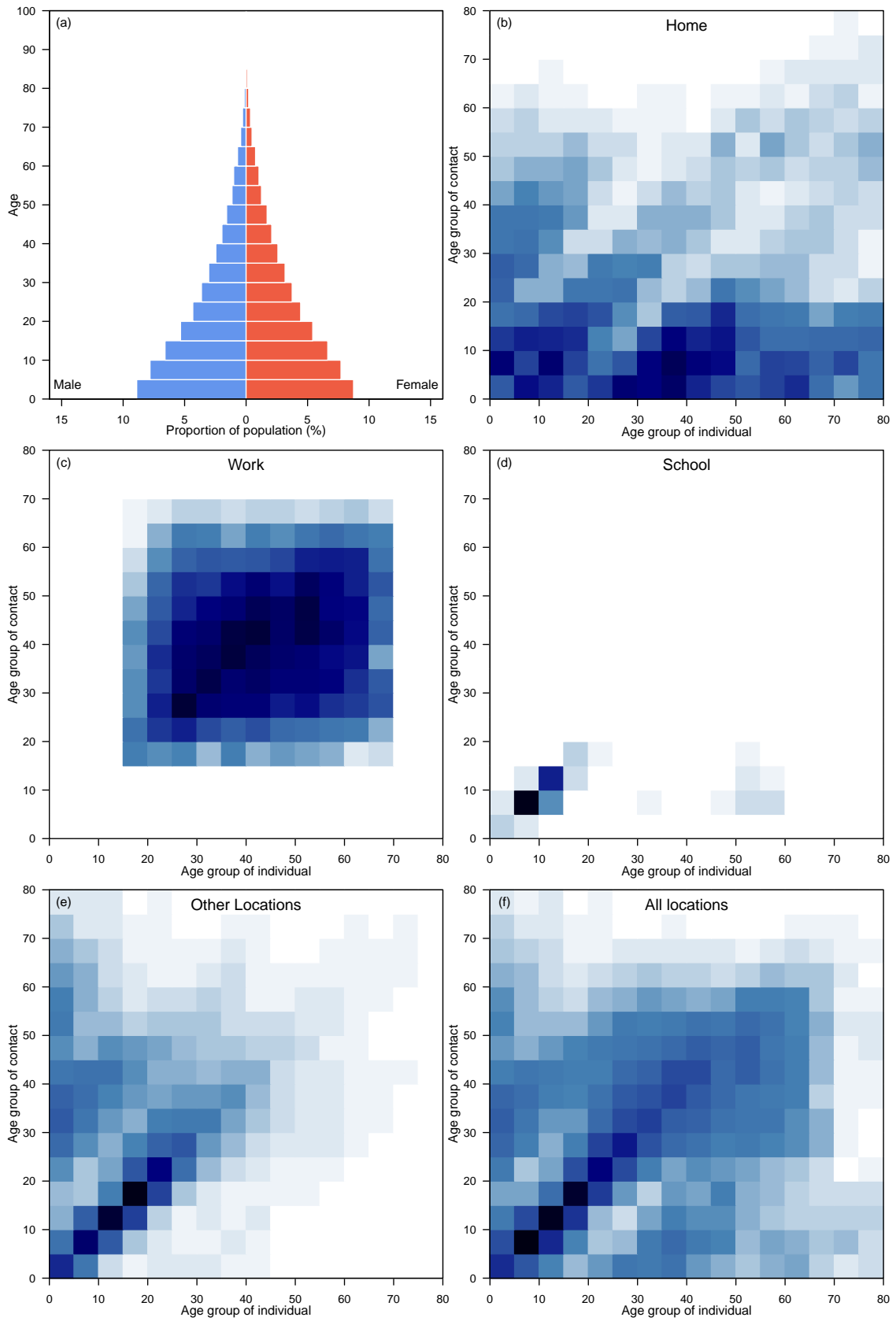
Argentina
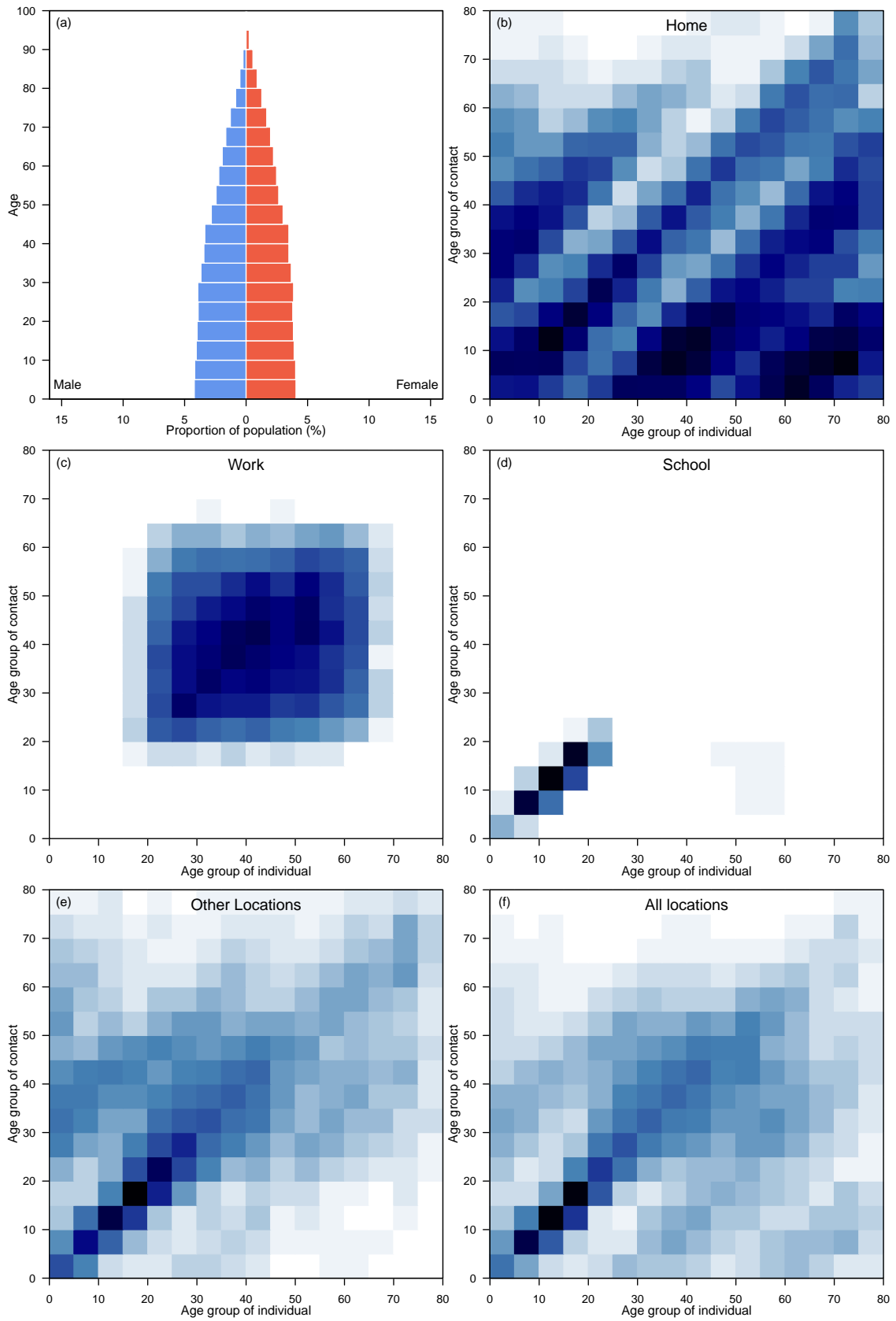
Armenia
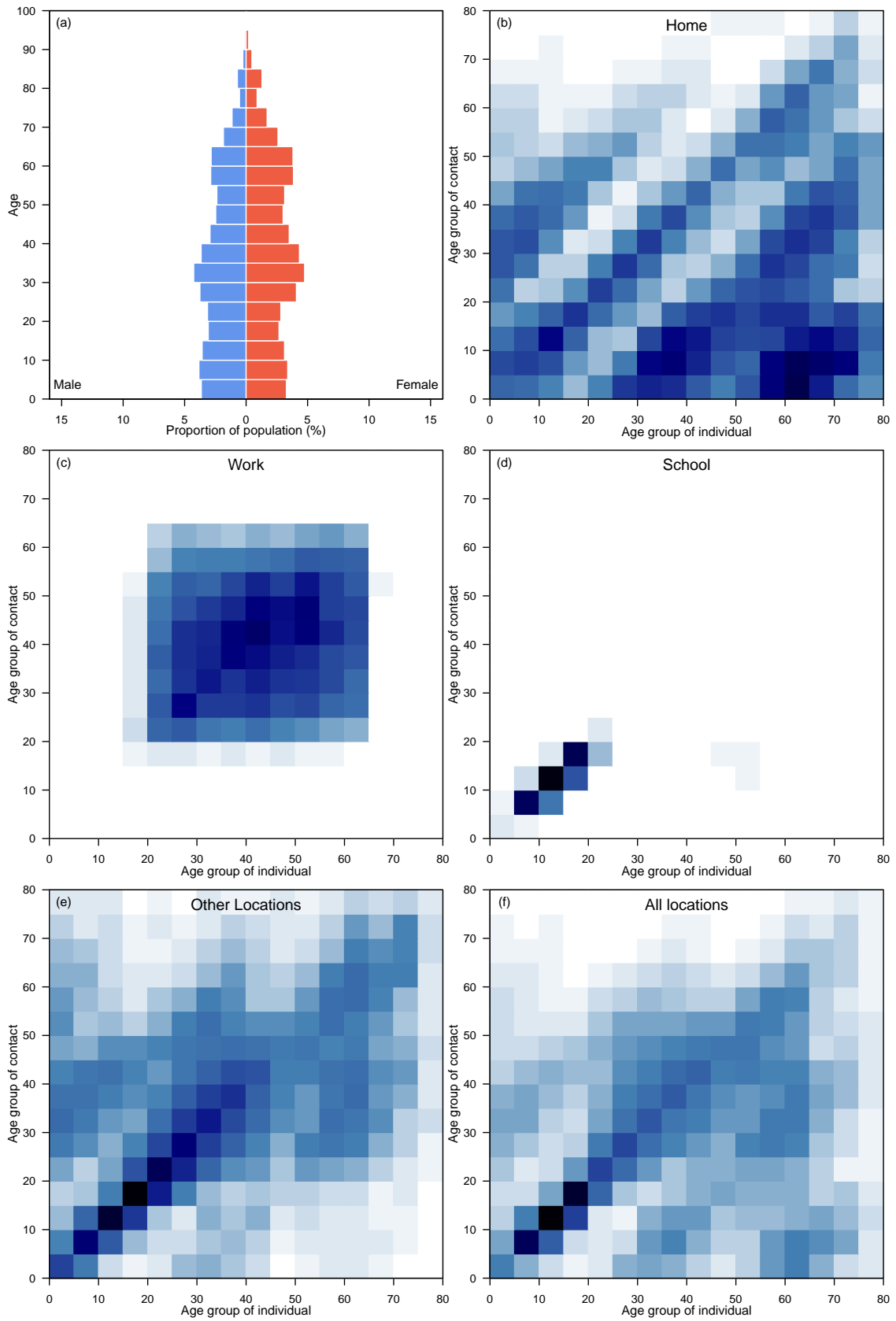
Austria
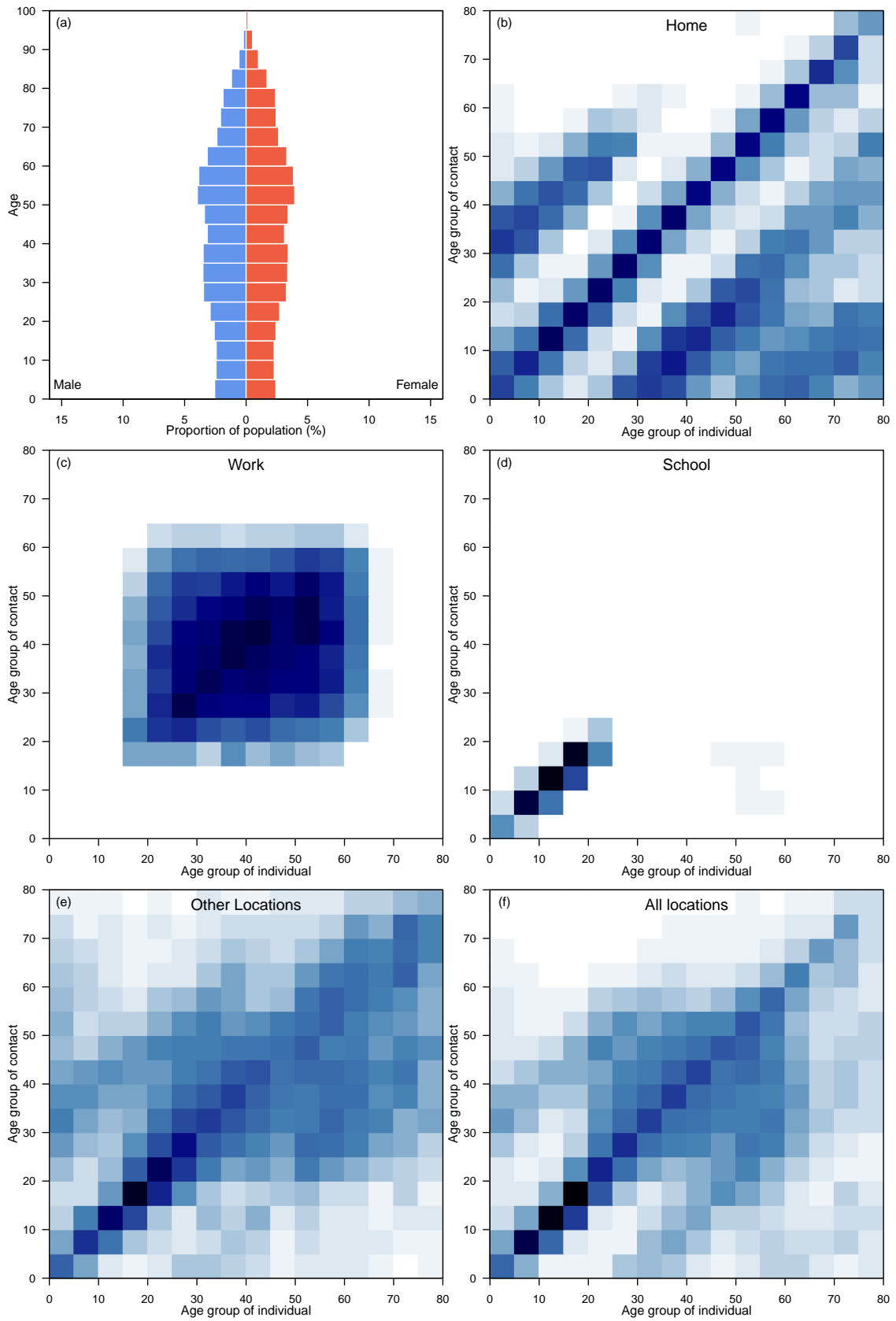
Azerbaijan
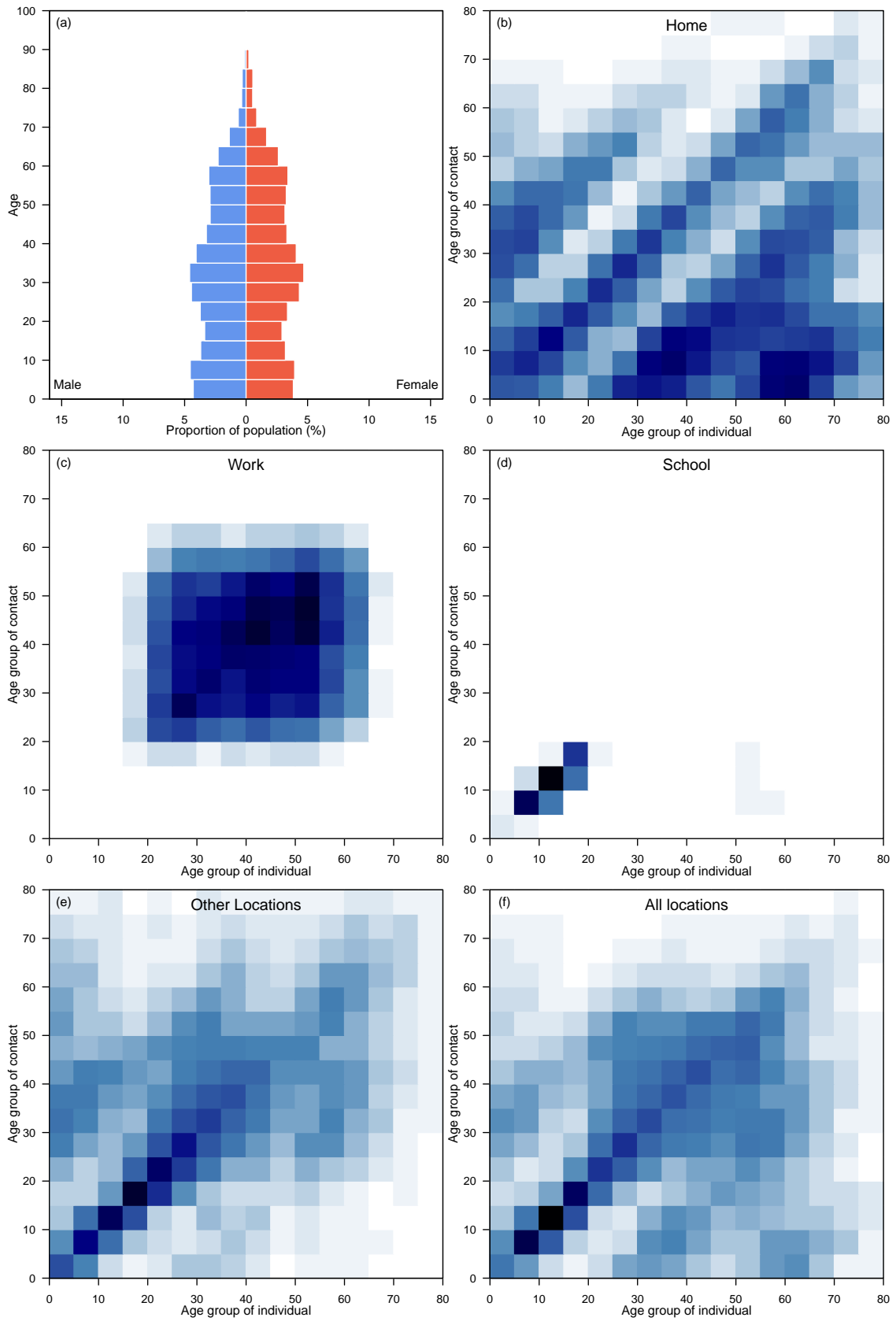
Bahamas
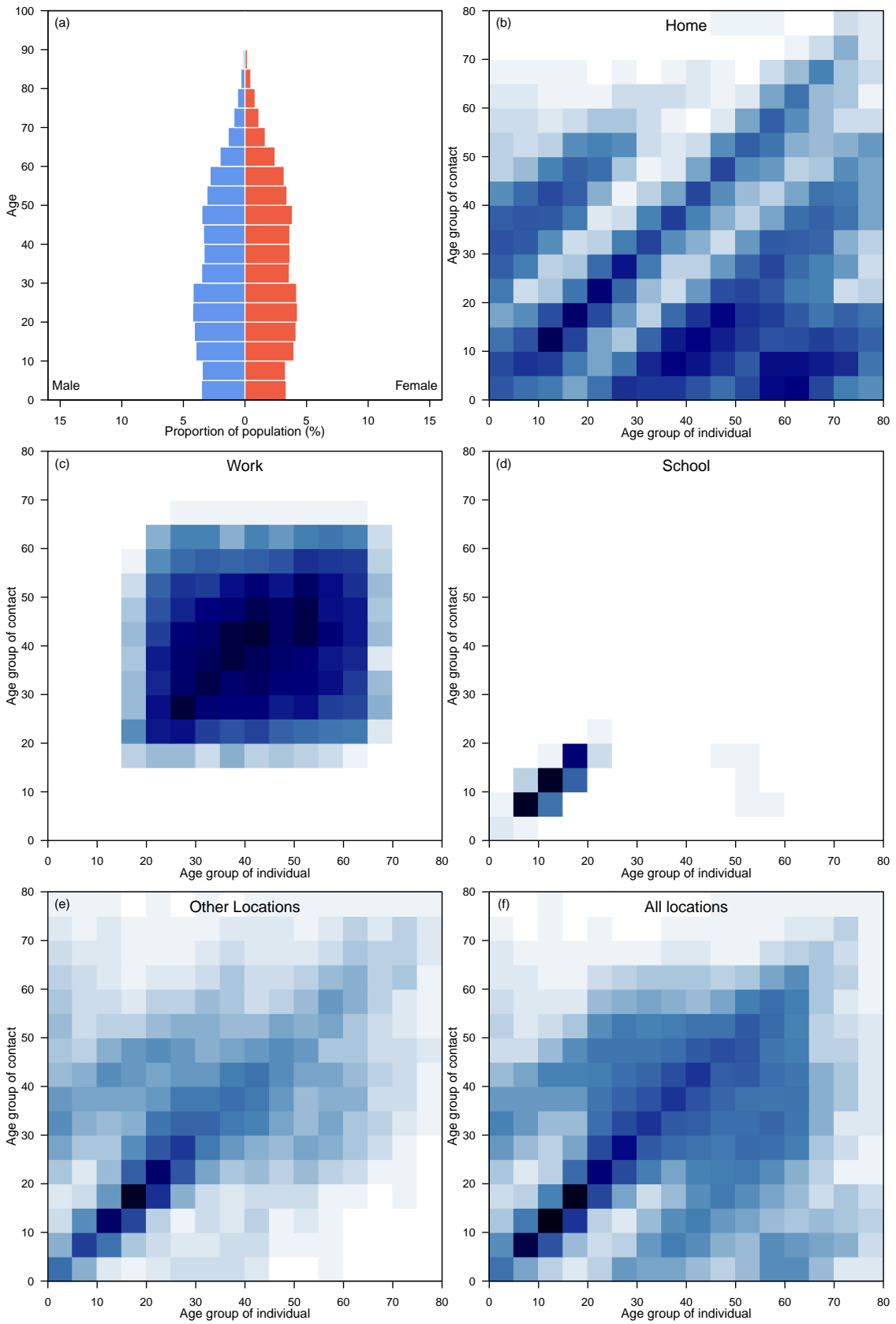
Bahrain
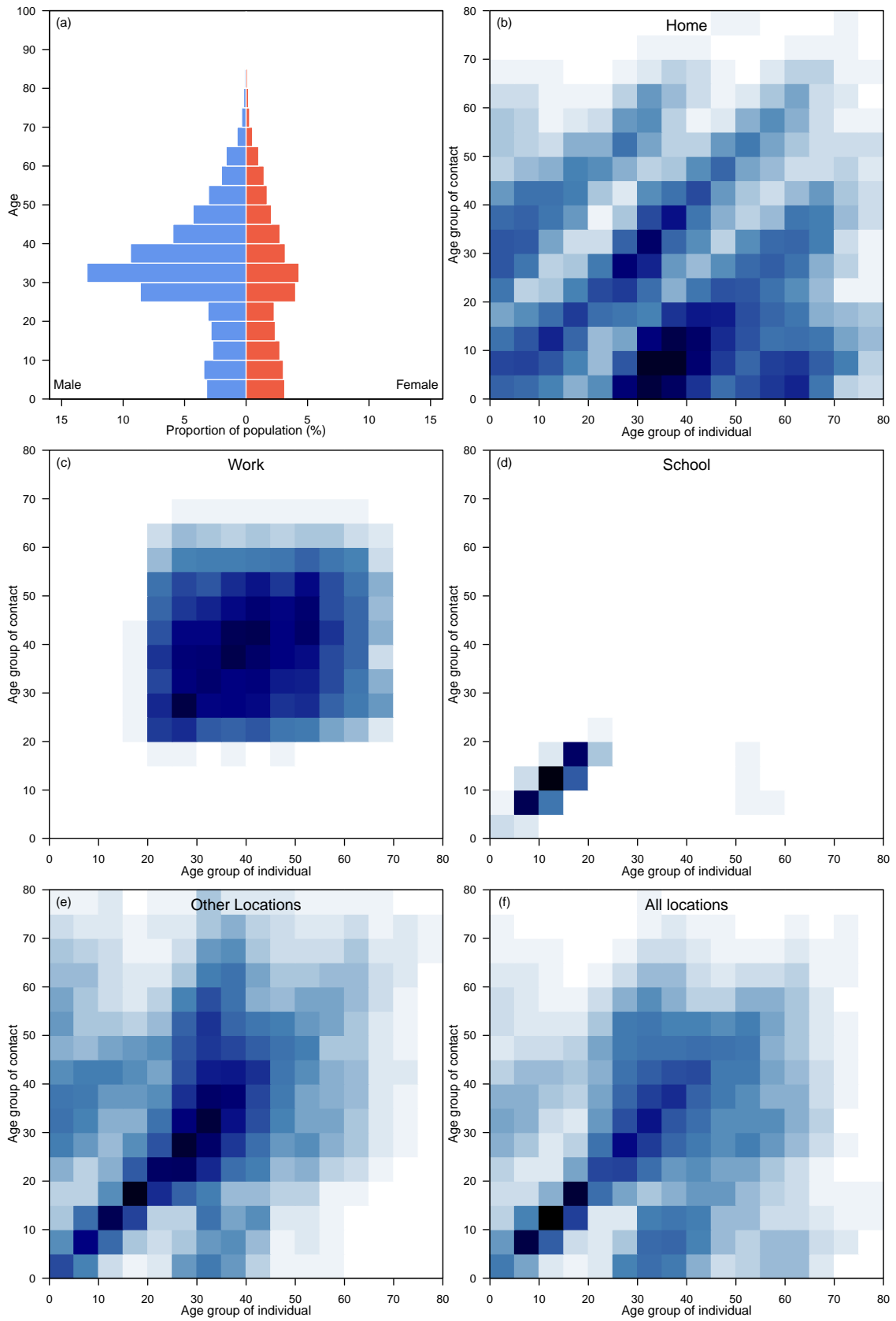
Bangladesh
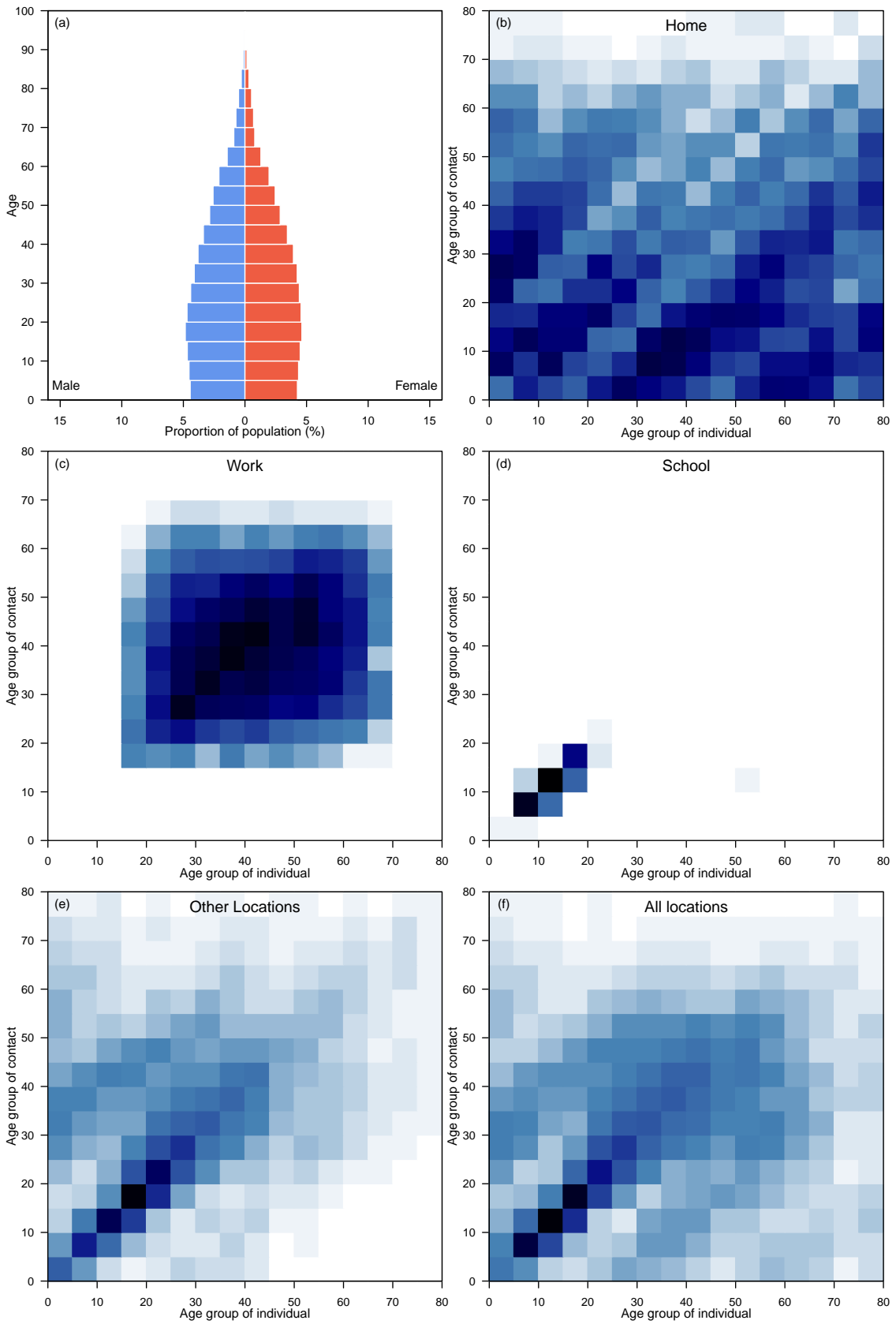
Barbados
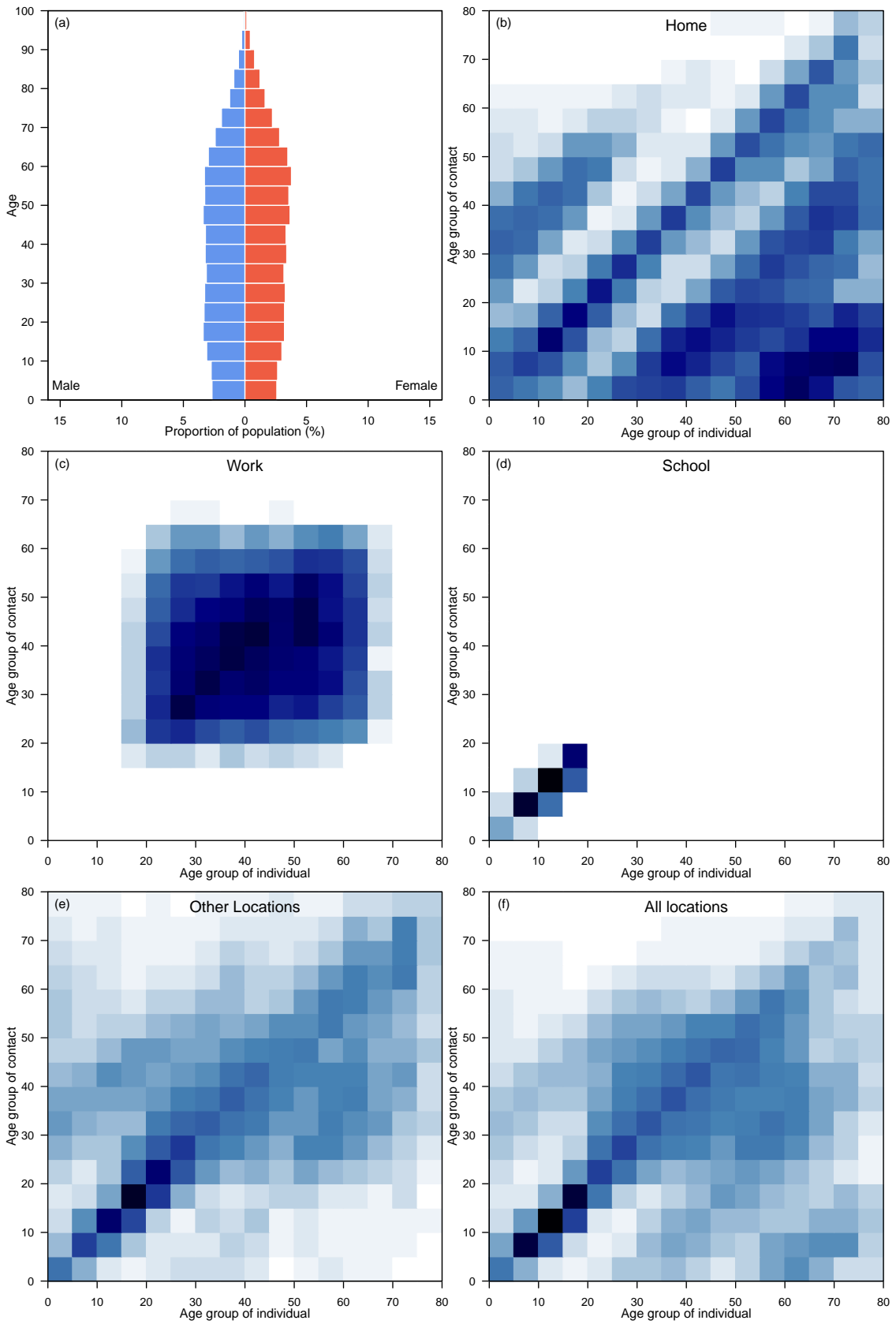
Belarus
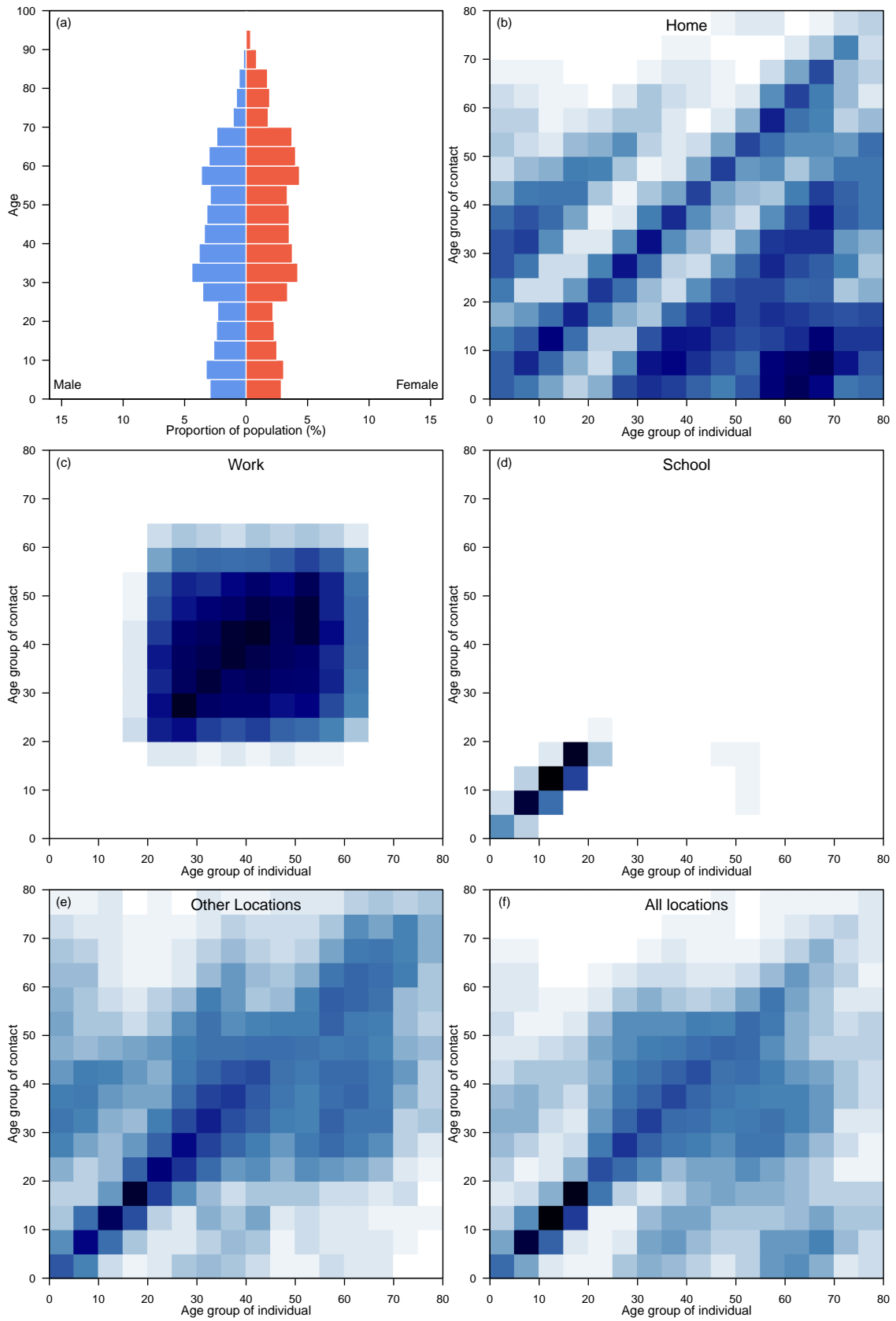
Belgium
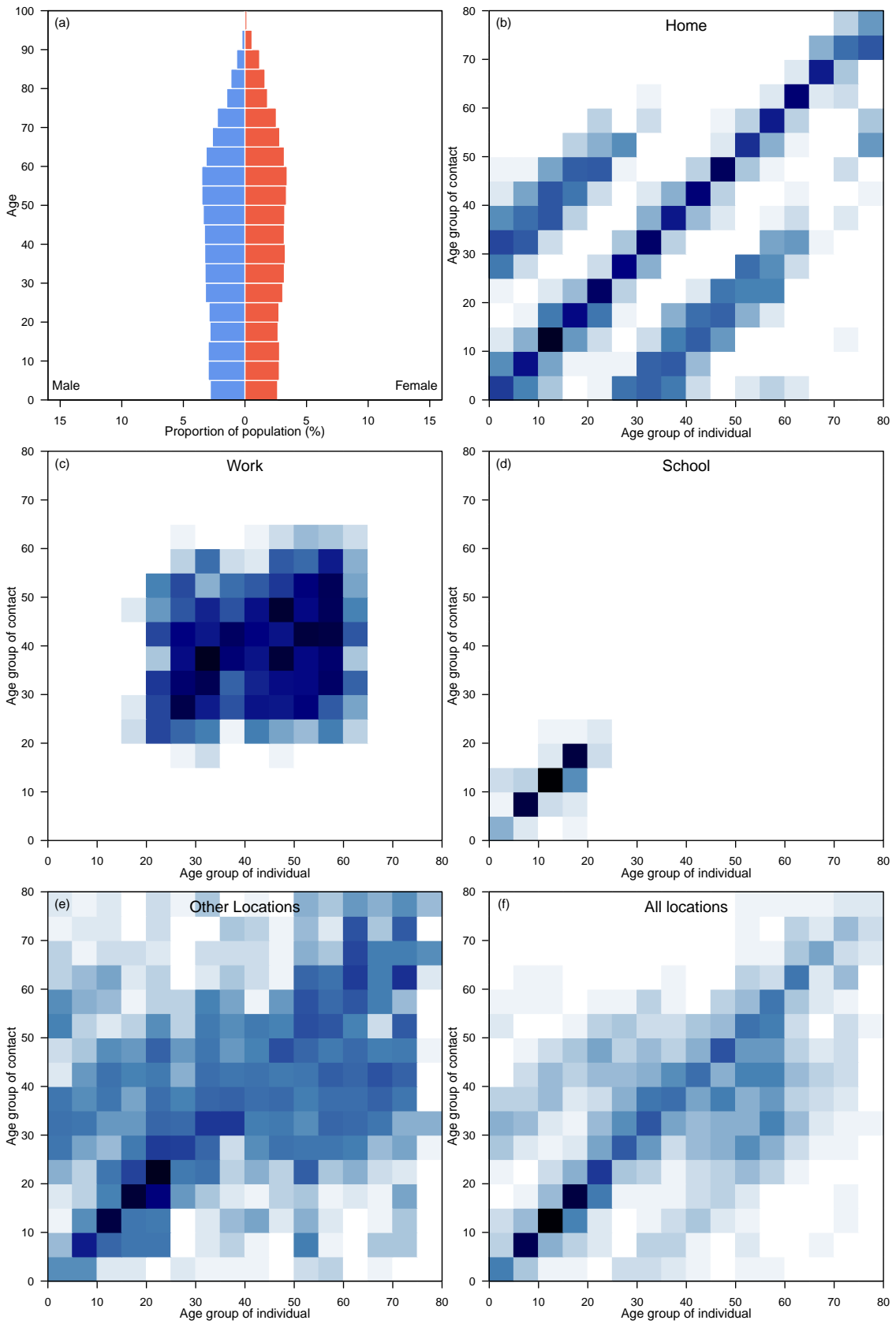
Belize
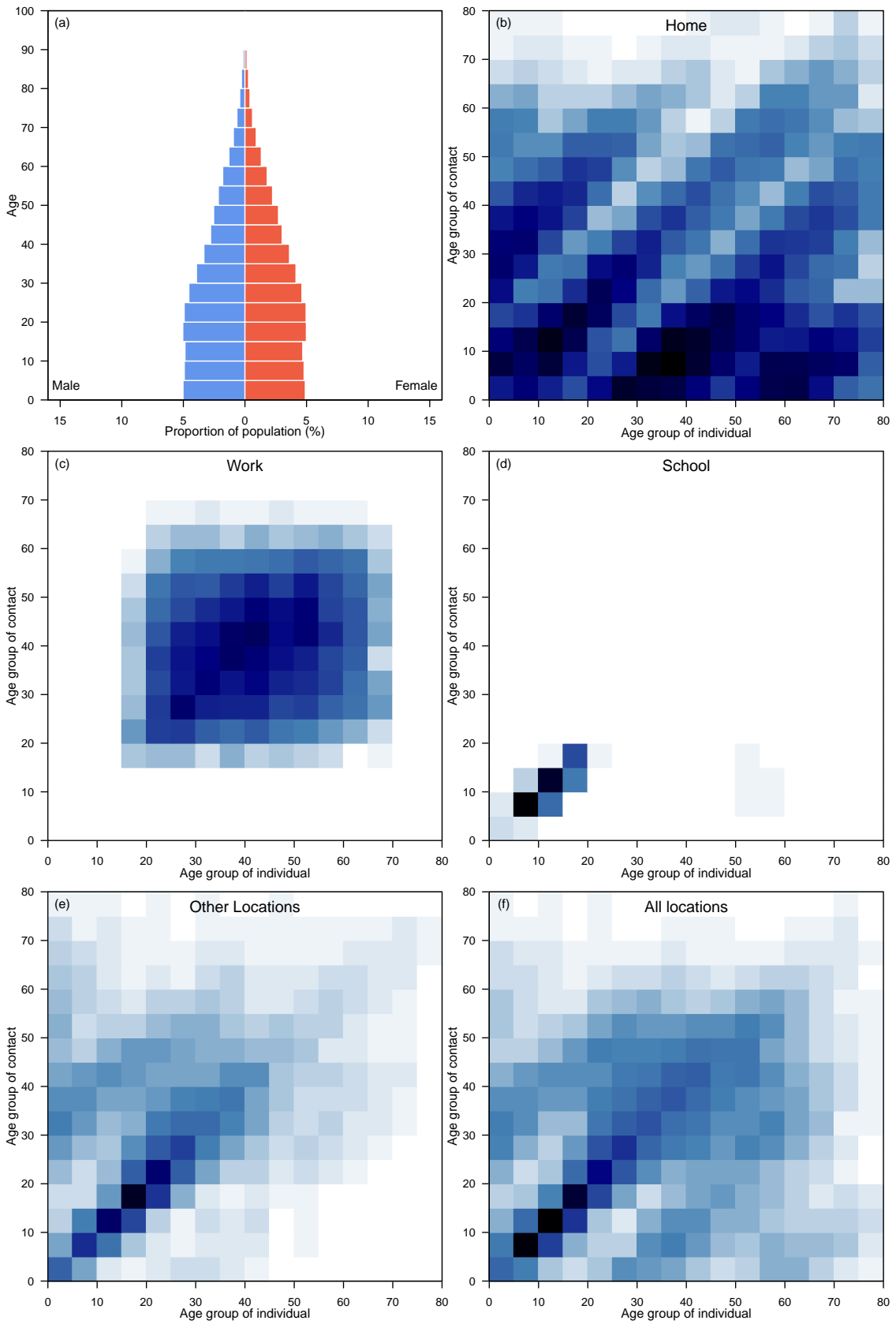
Benin
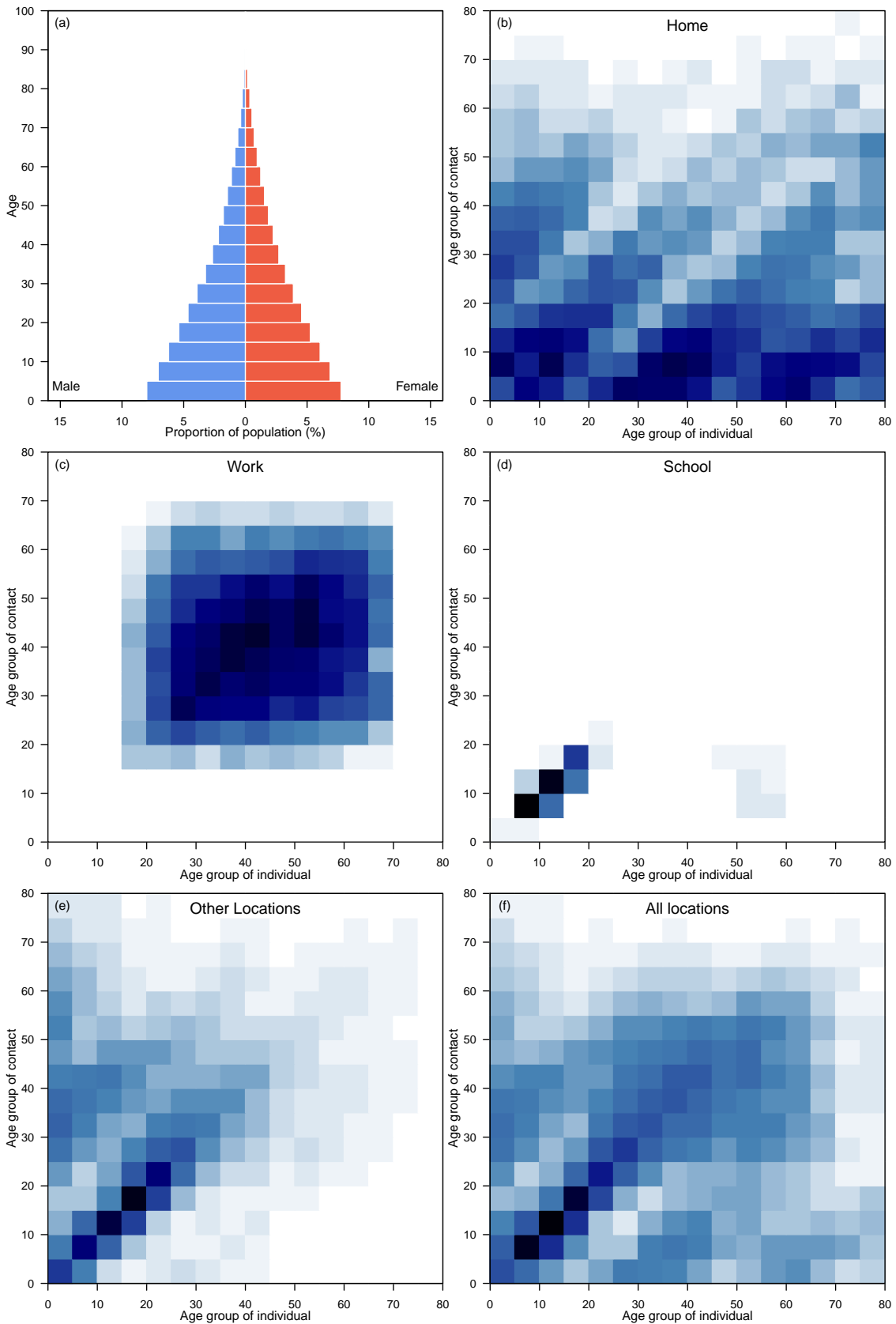
Bhutan
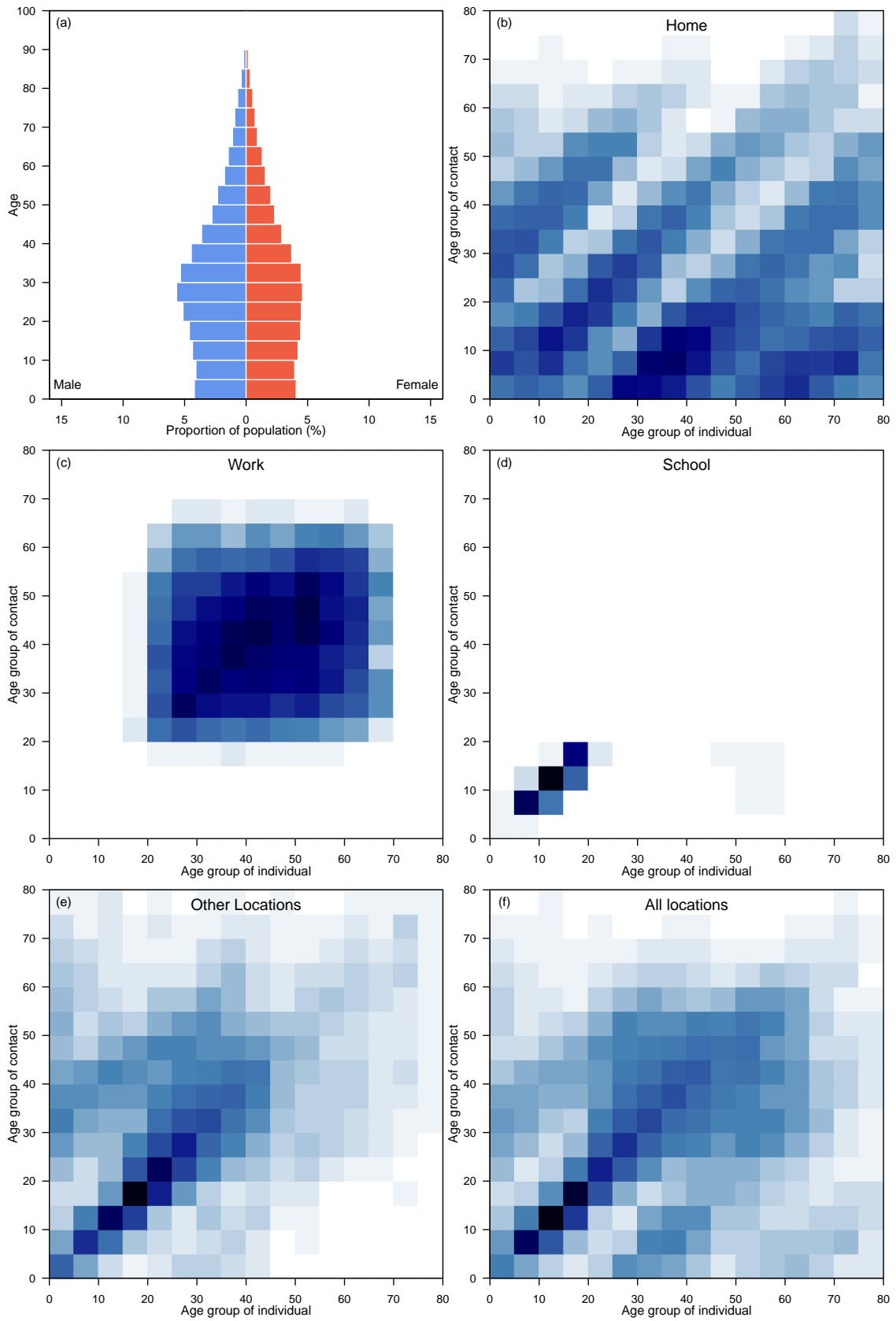
Bolivia (Plurinational State of)
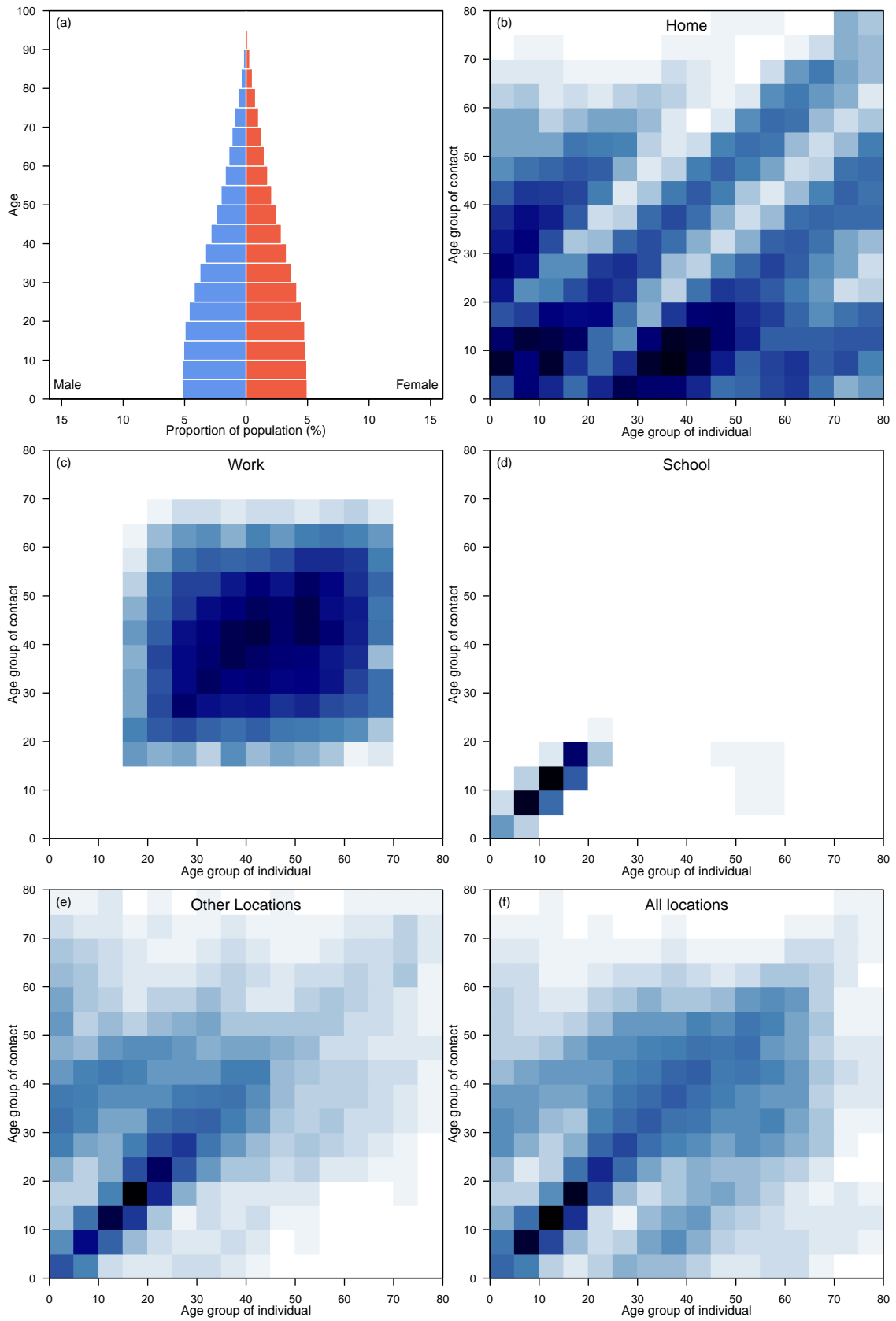
Bosnia and Herzegovina
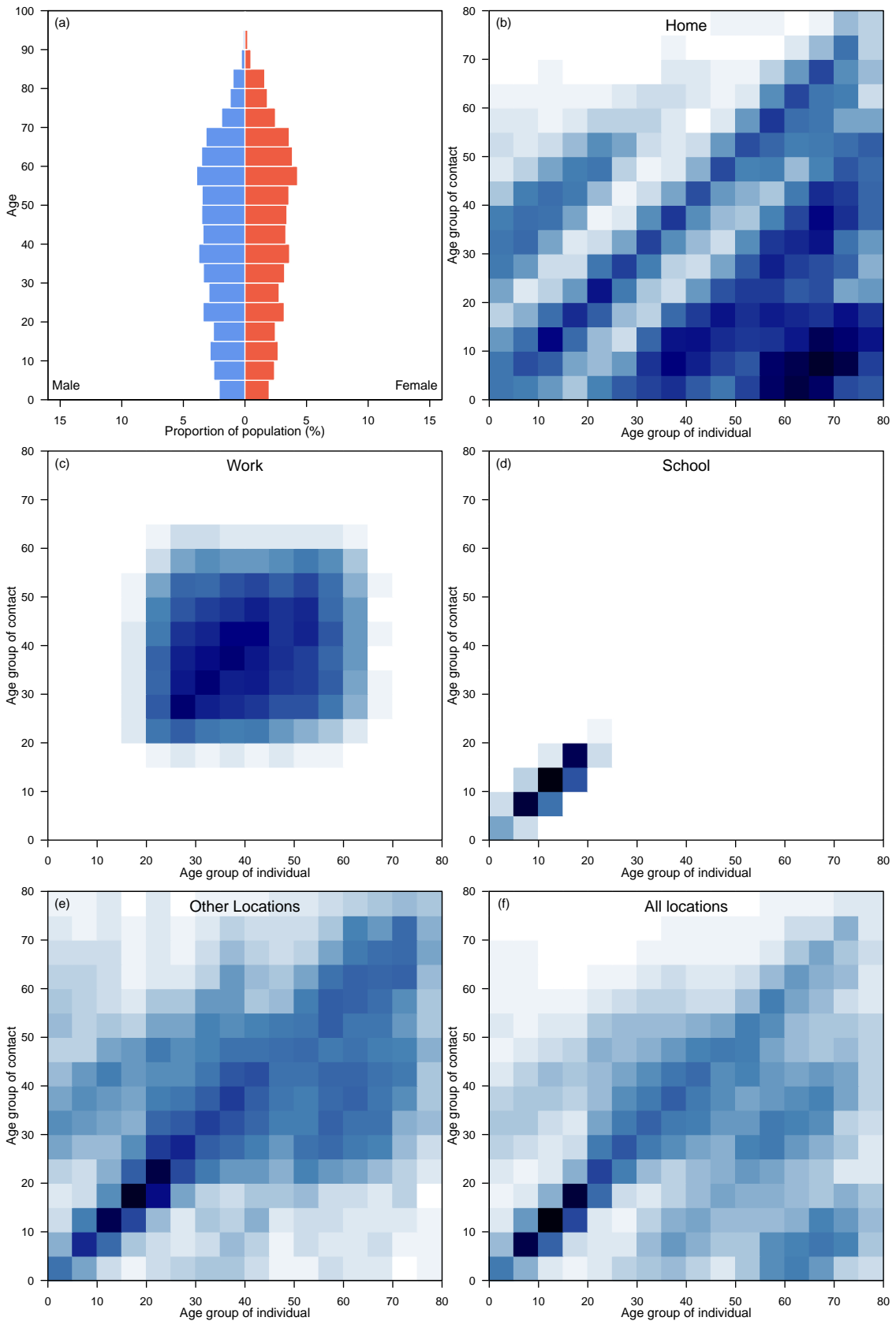
Botswana
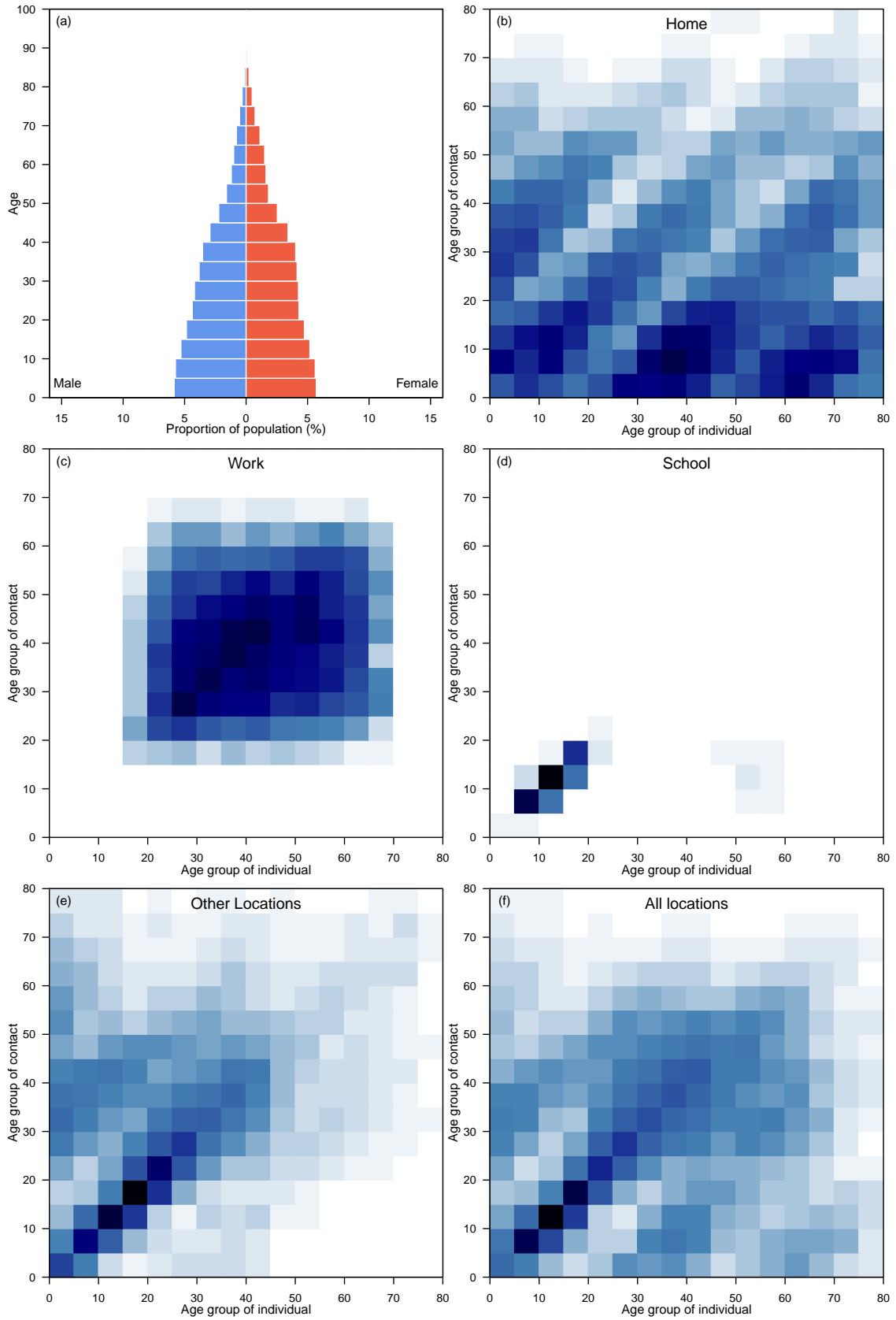
Brazil
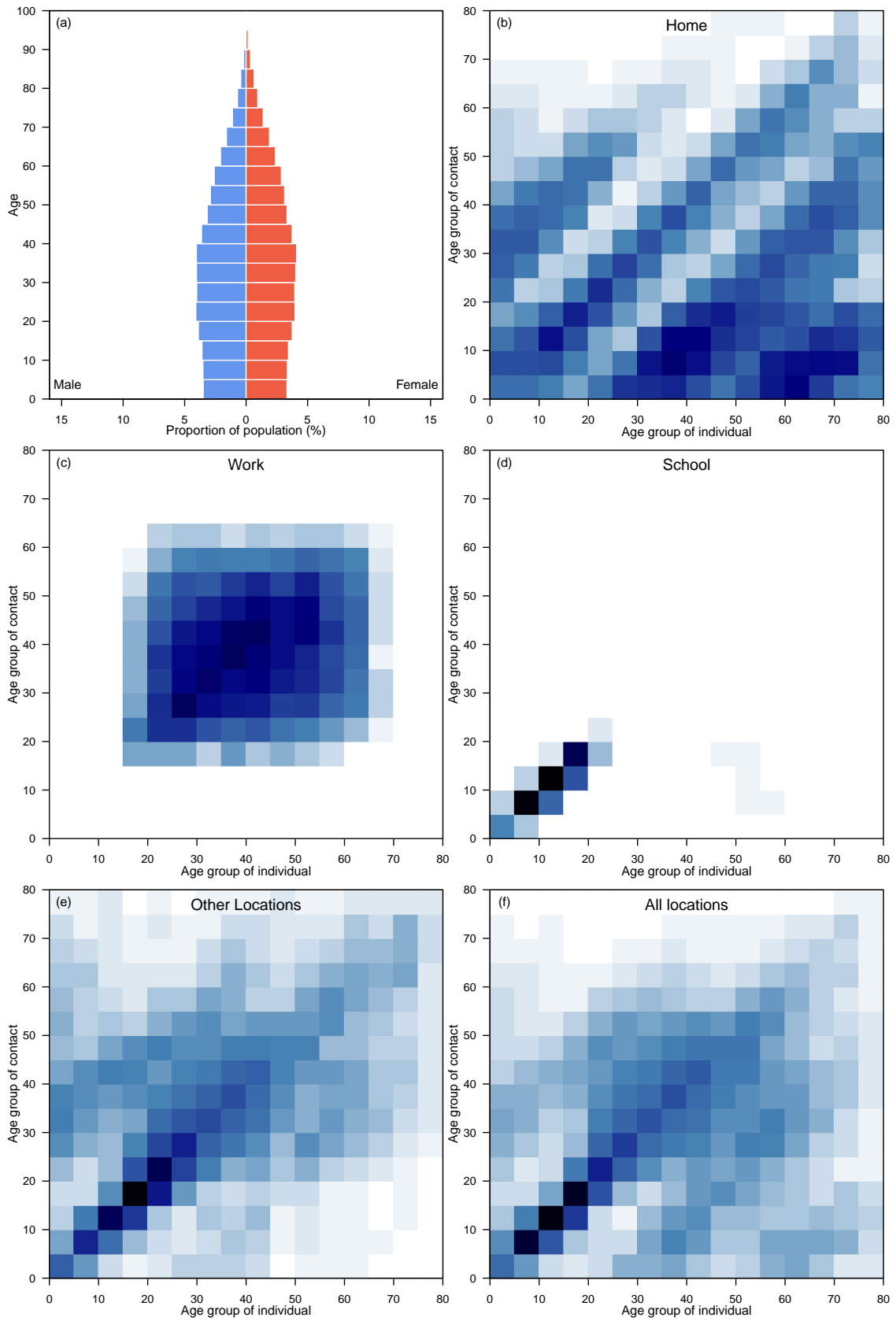
Brunei Darussalam
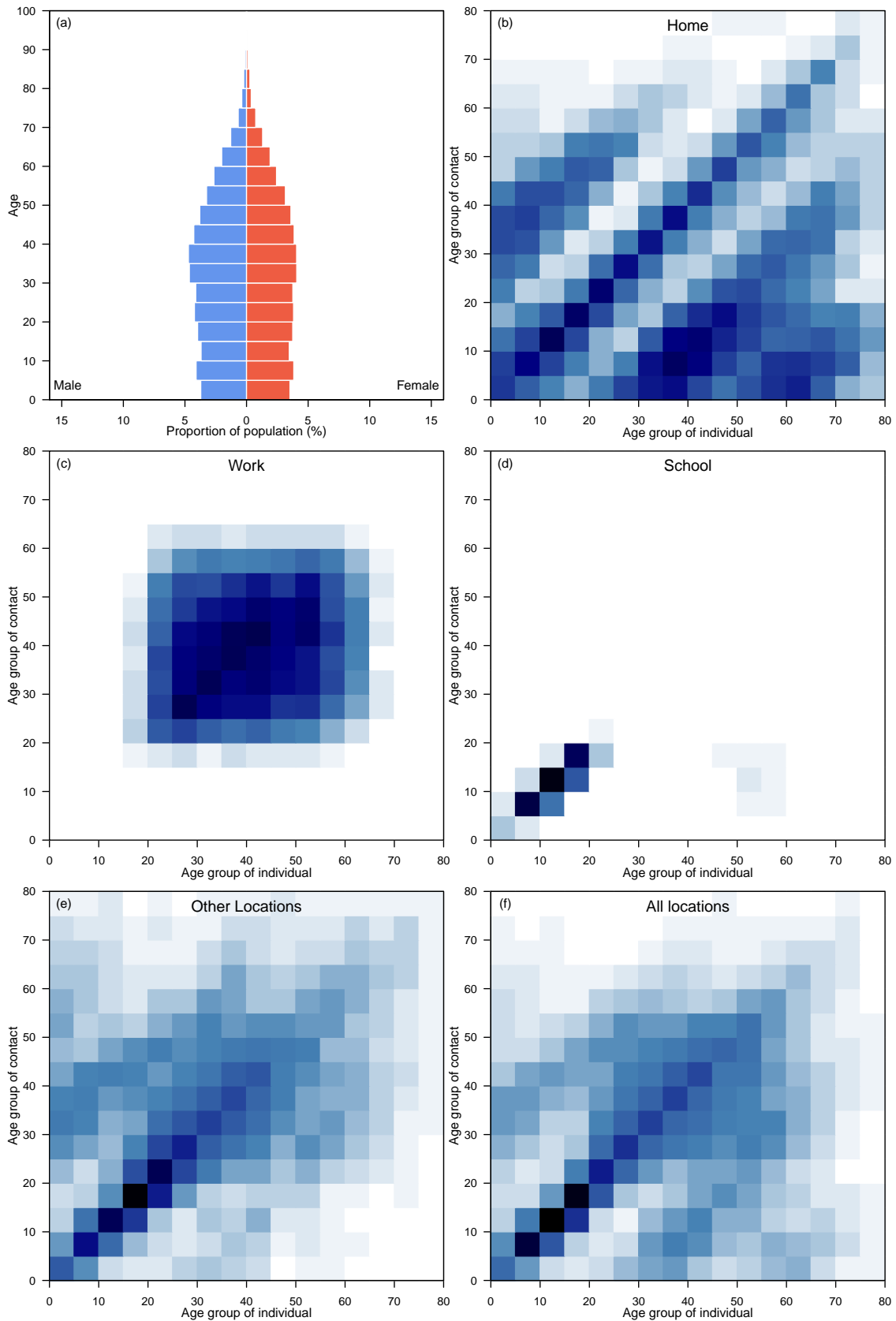
Bulgaria
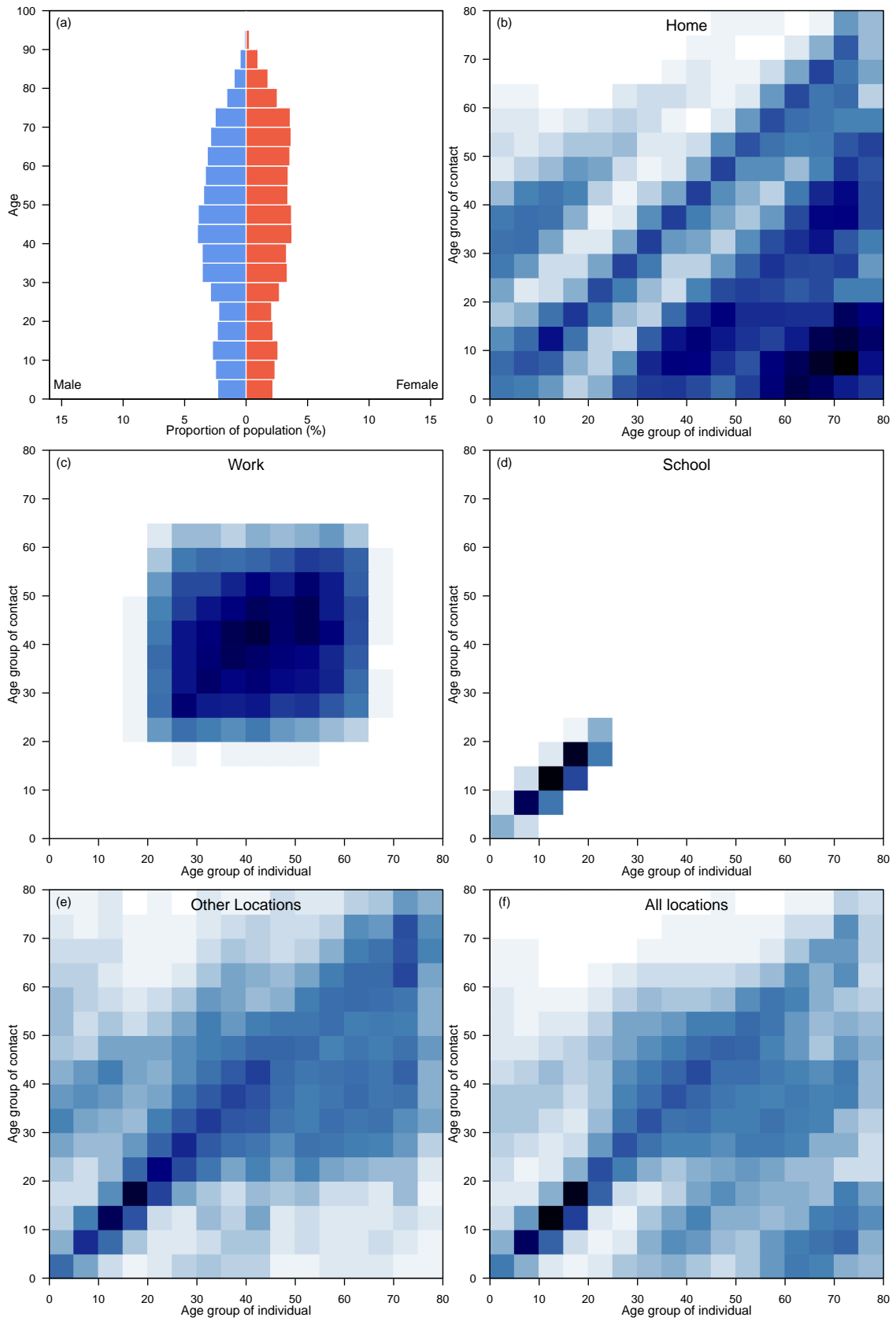
Burkina Faso
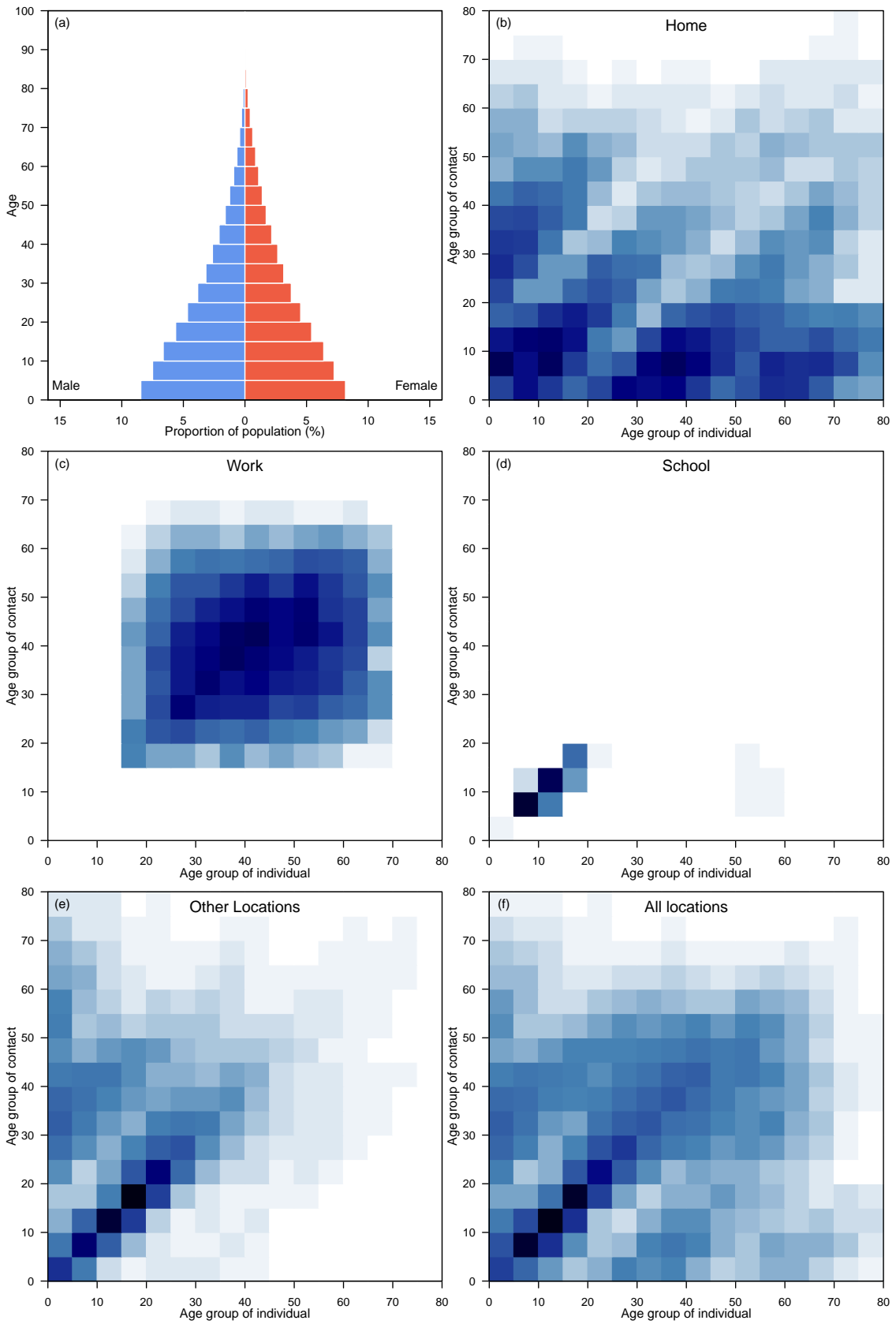
Burundi
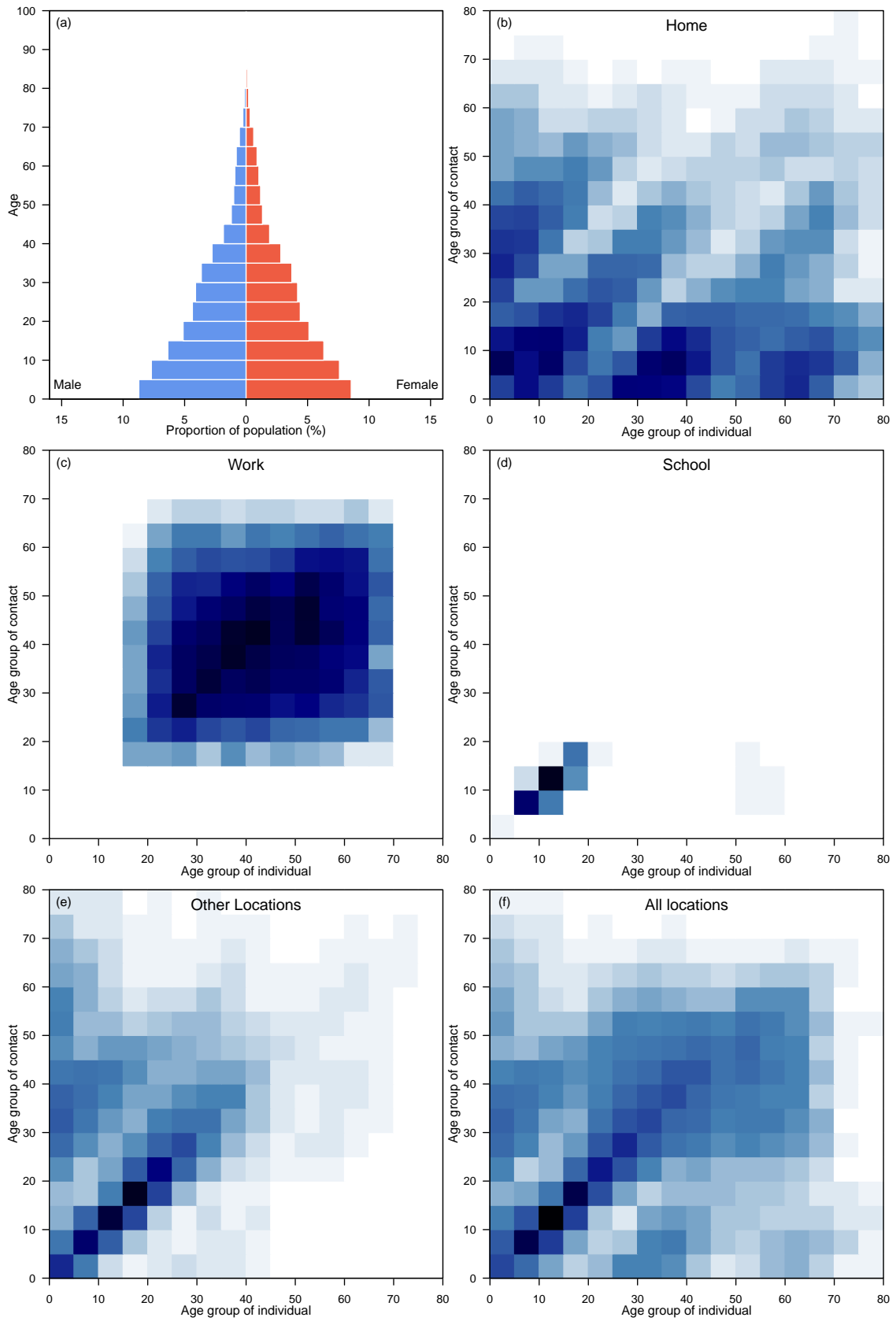
Cabo Verde
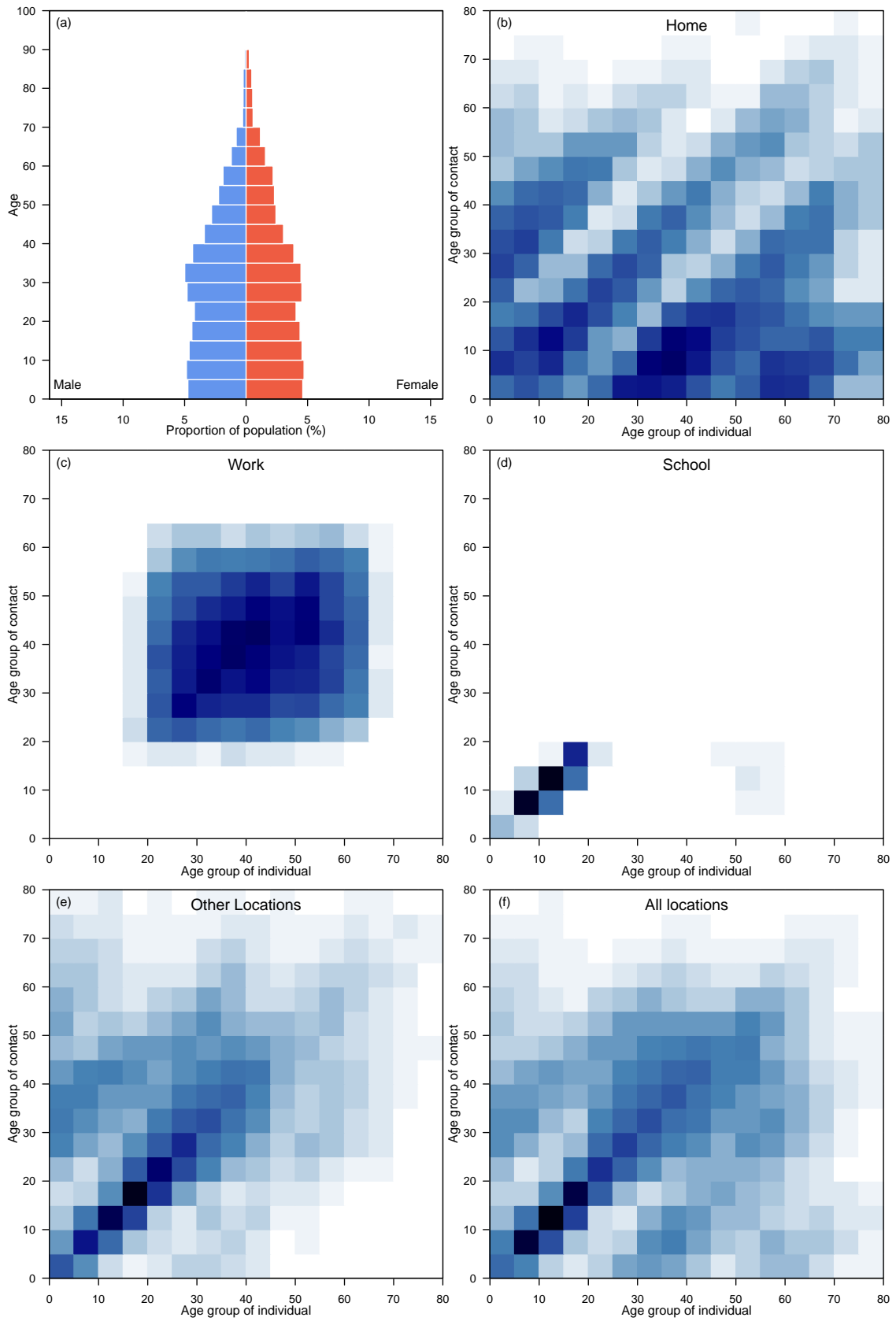
Cambodia
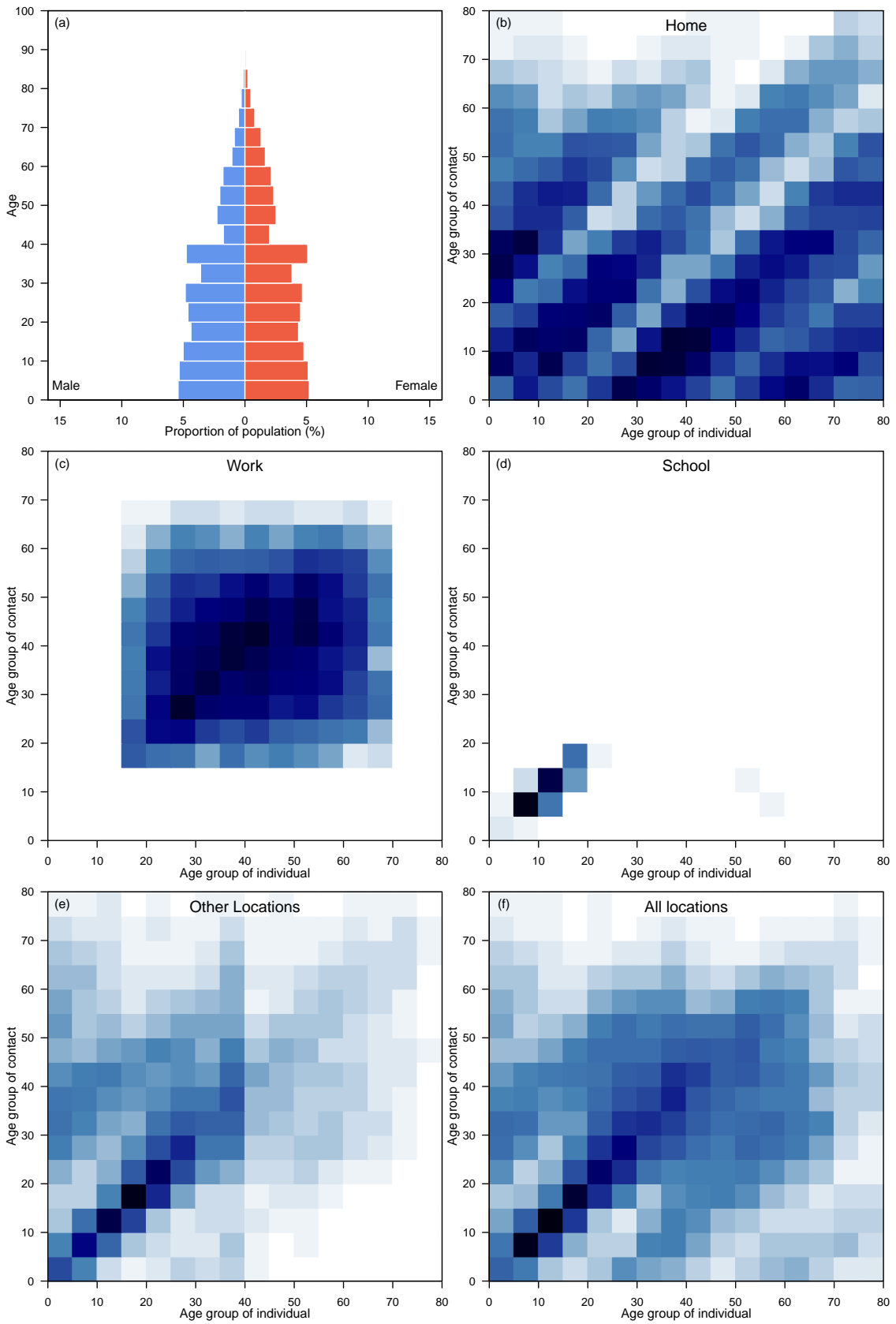
Cameroon
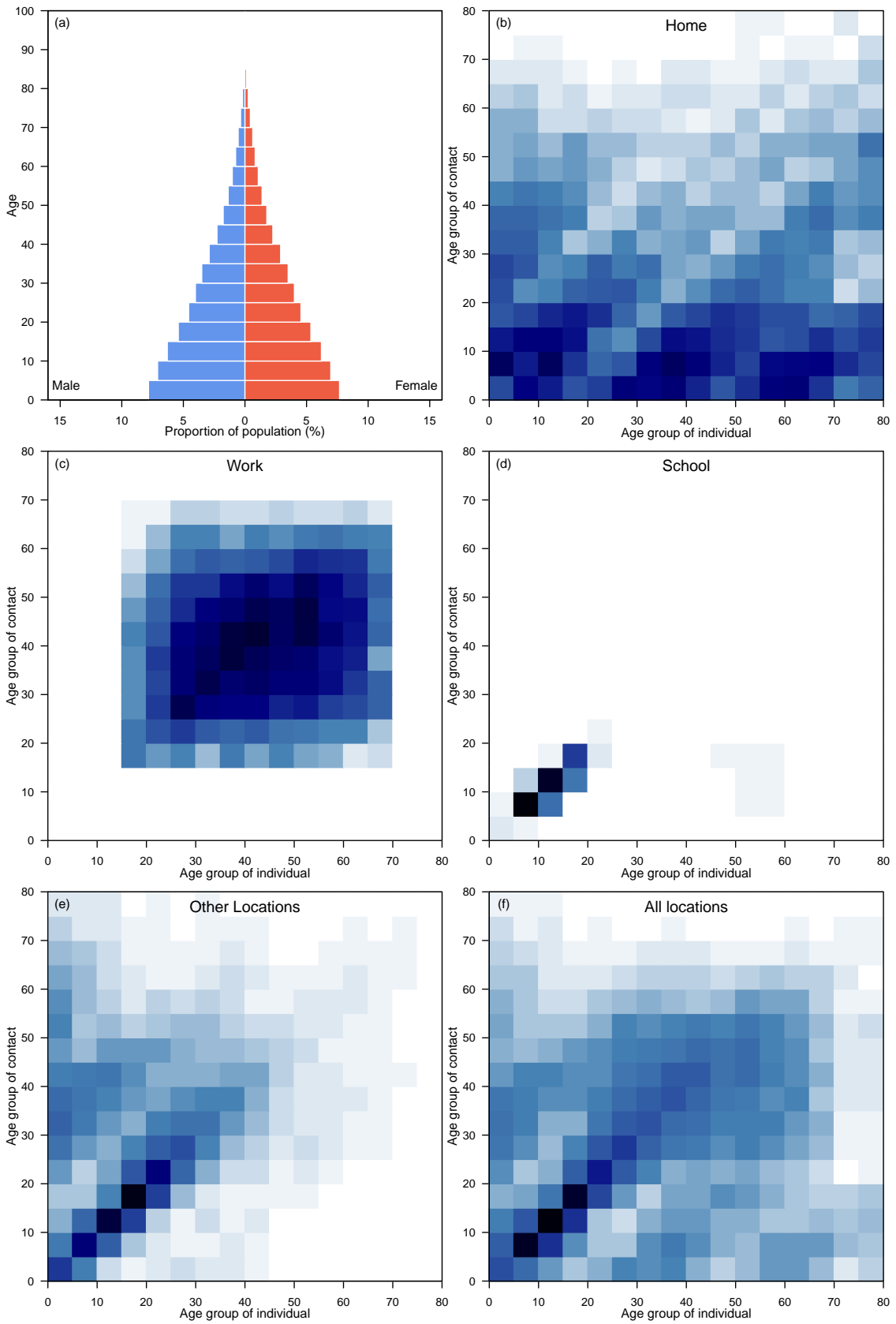
Canada
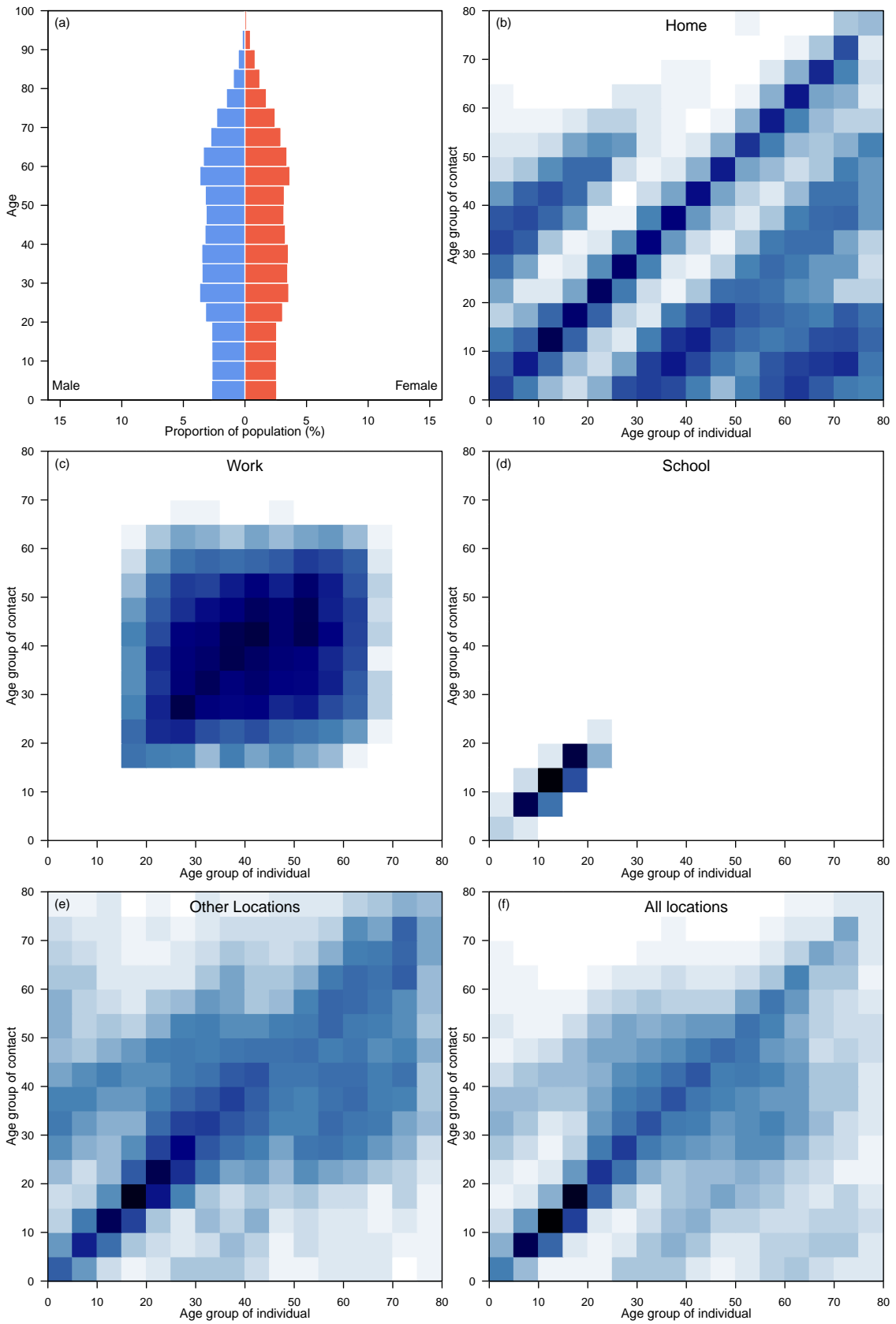
Central African Republic
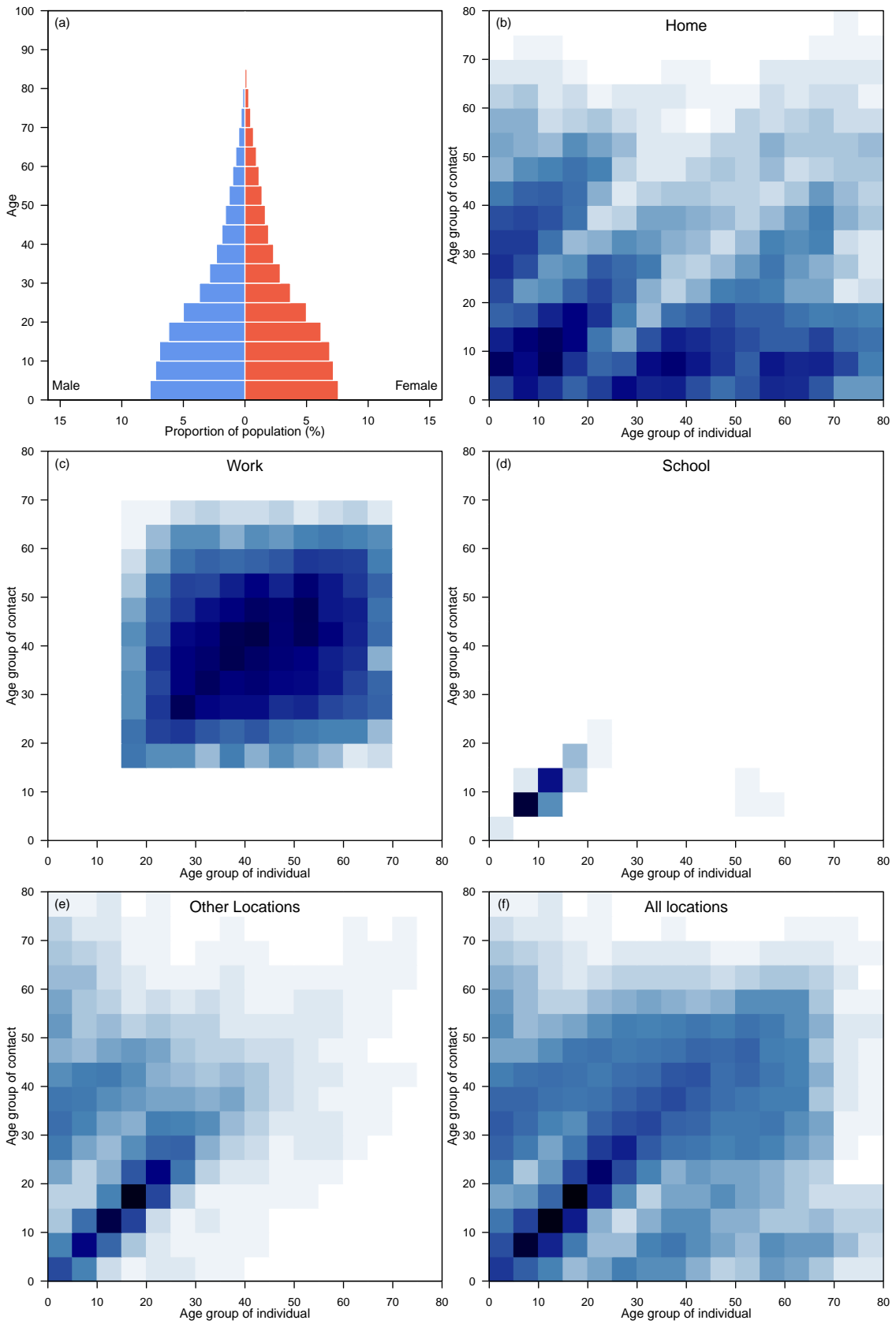
Chad
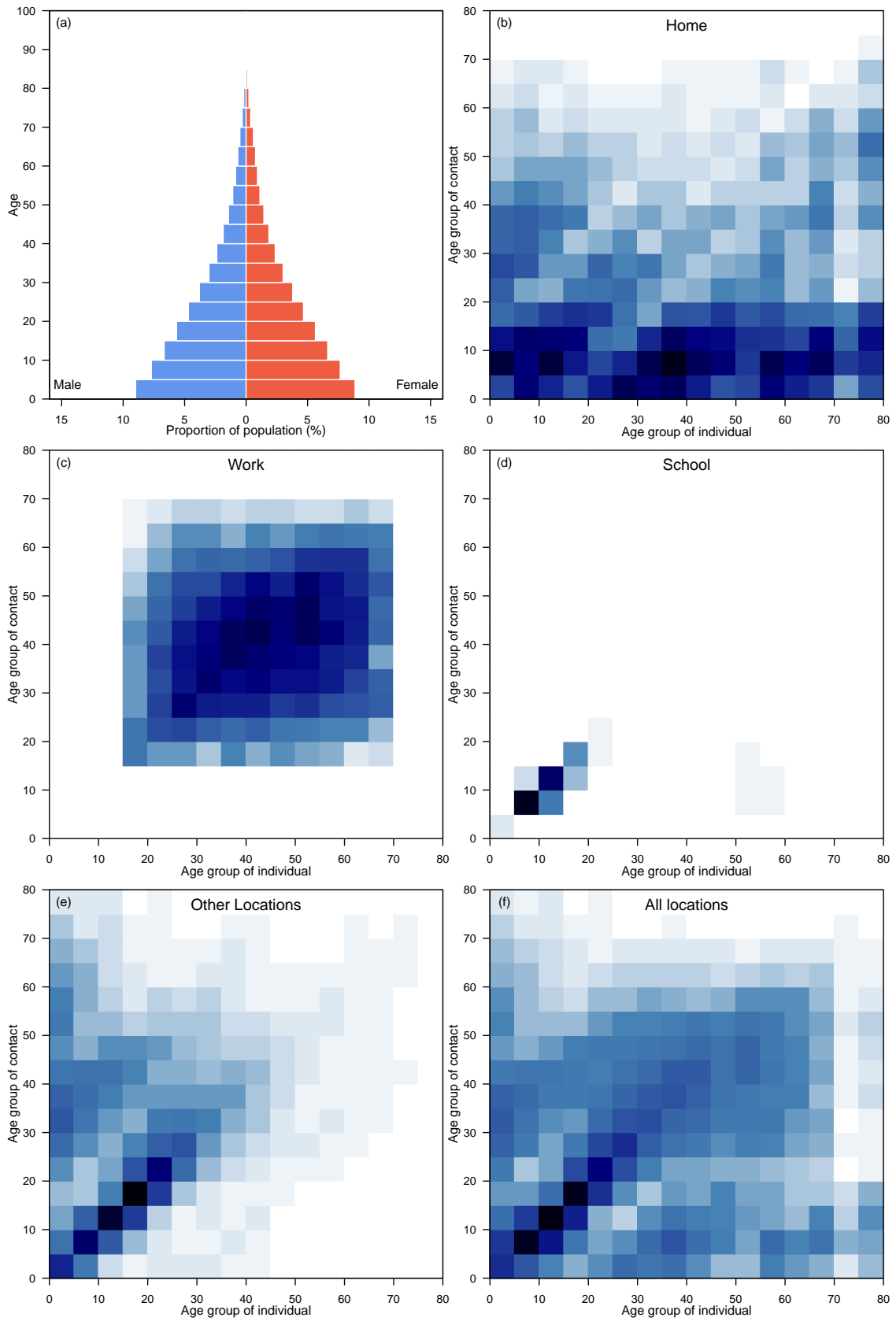
Chile
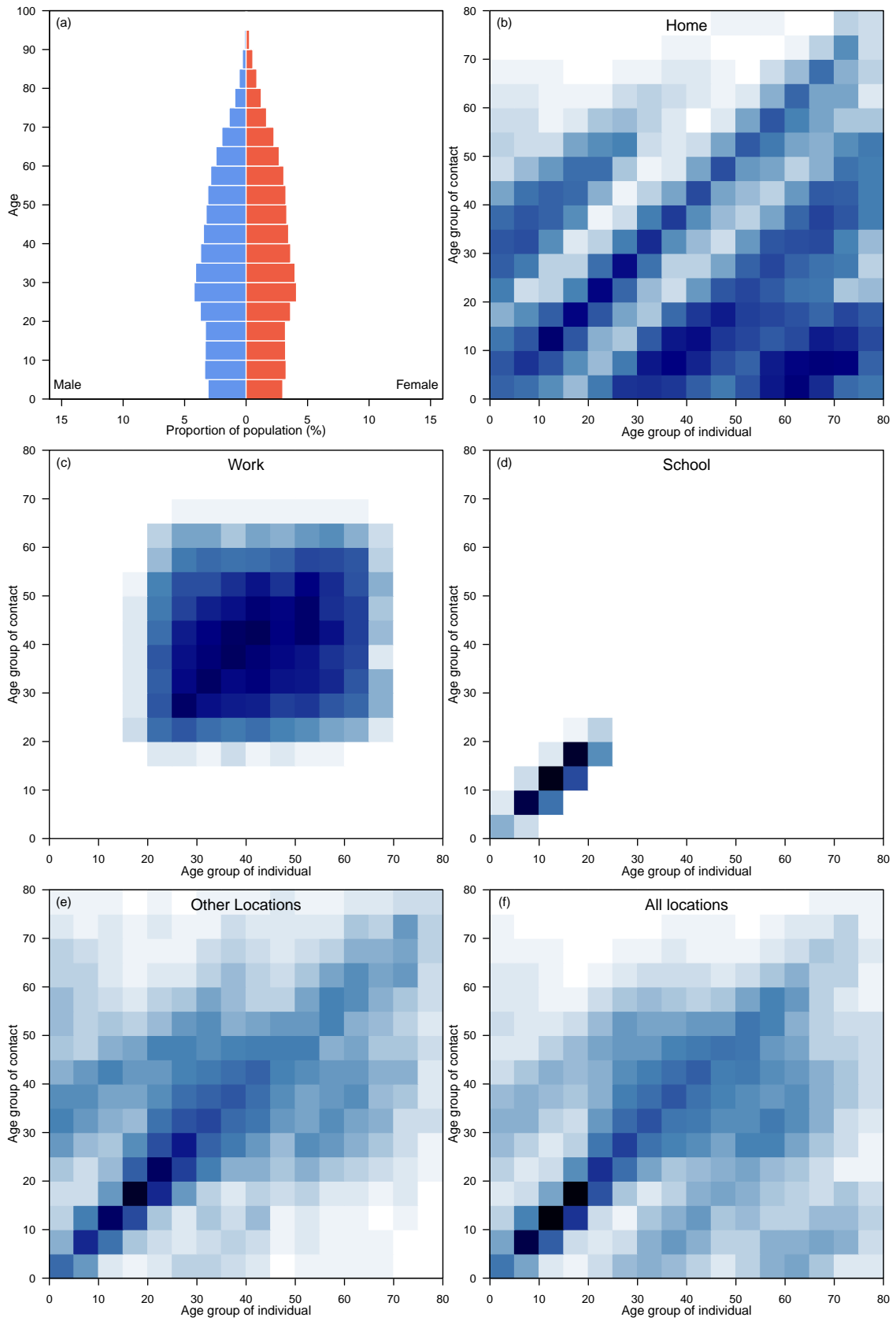
China
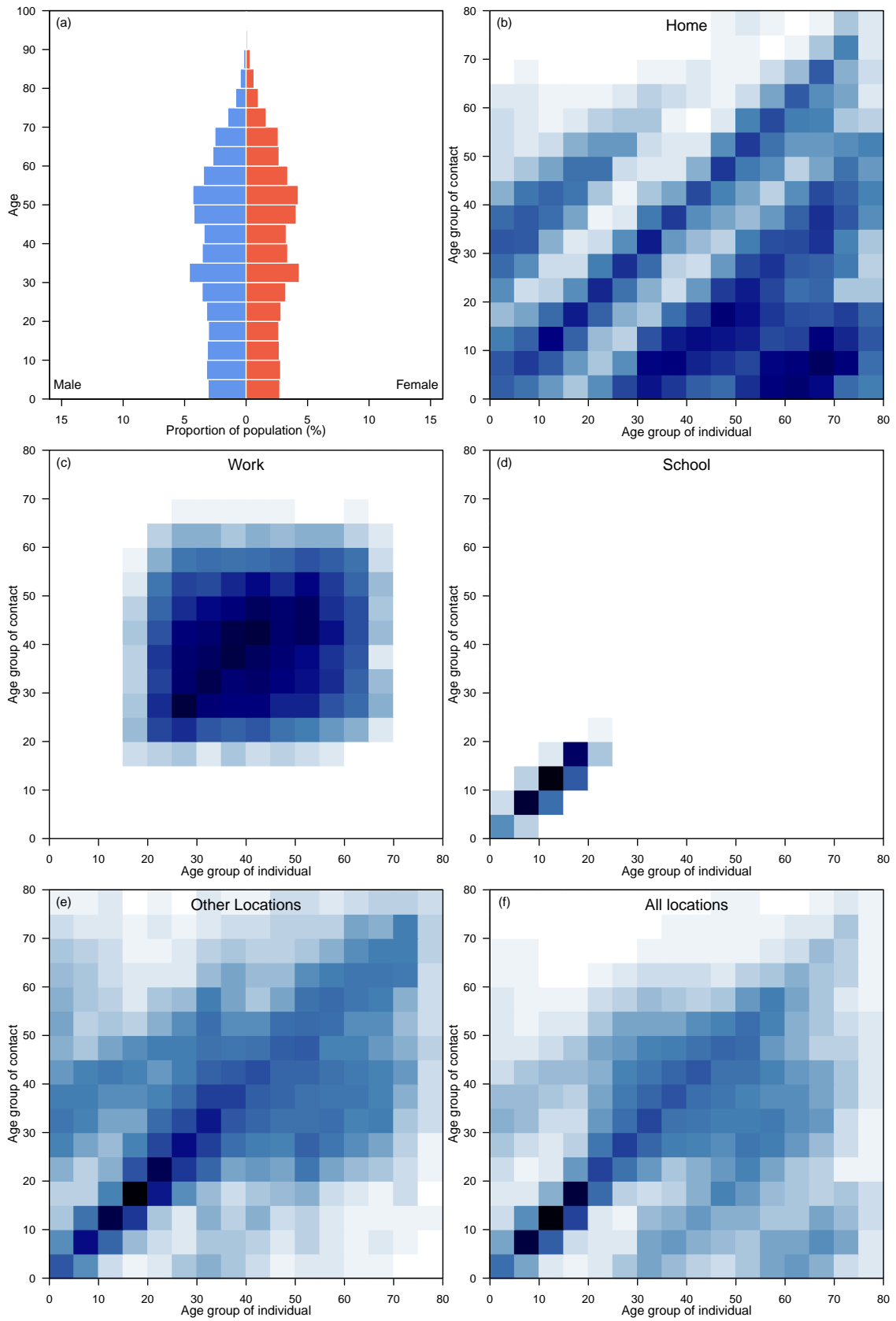
China, Hong Kong SAR
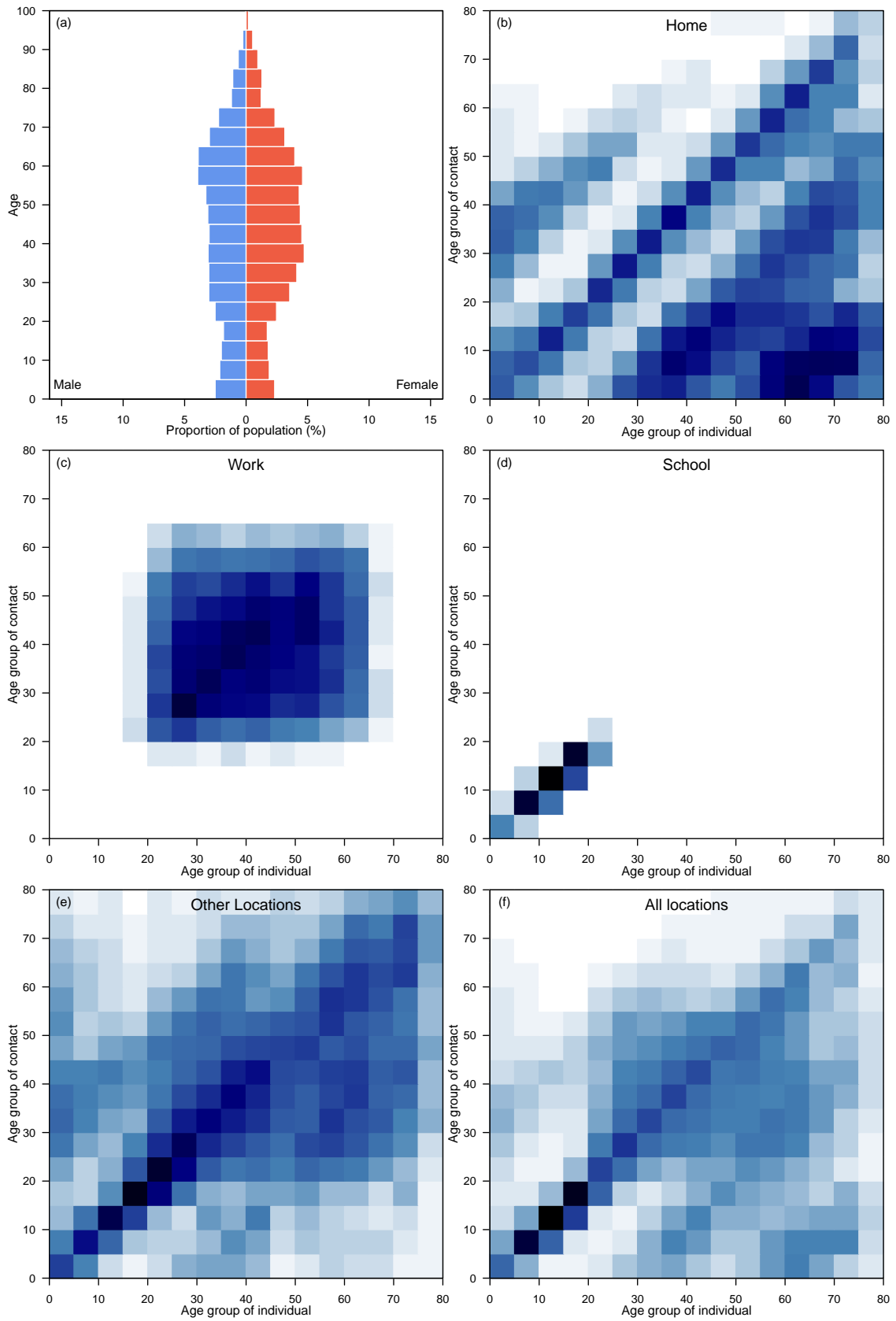
China, Macao SAR
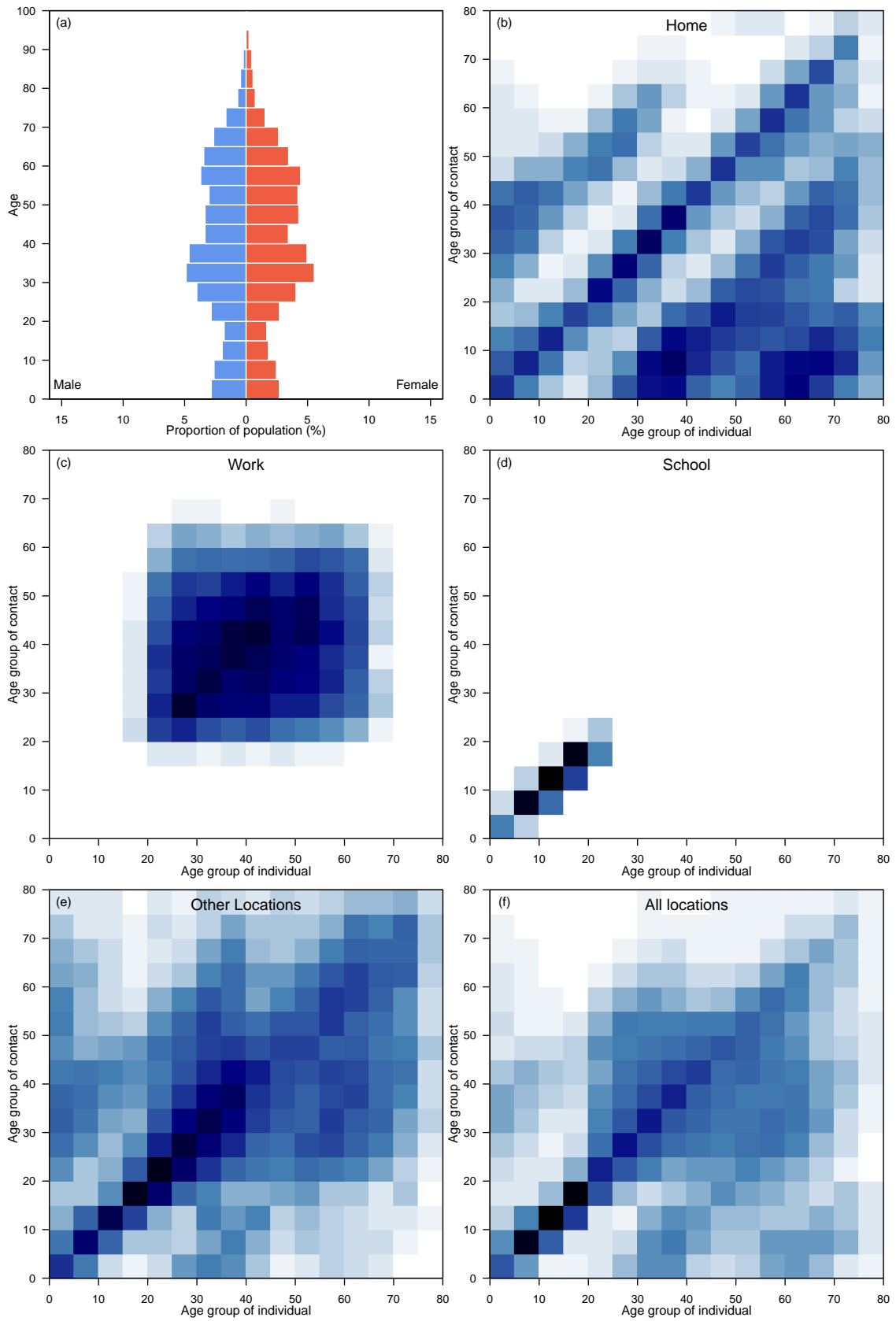
Colombia
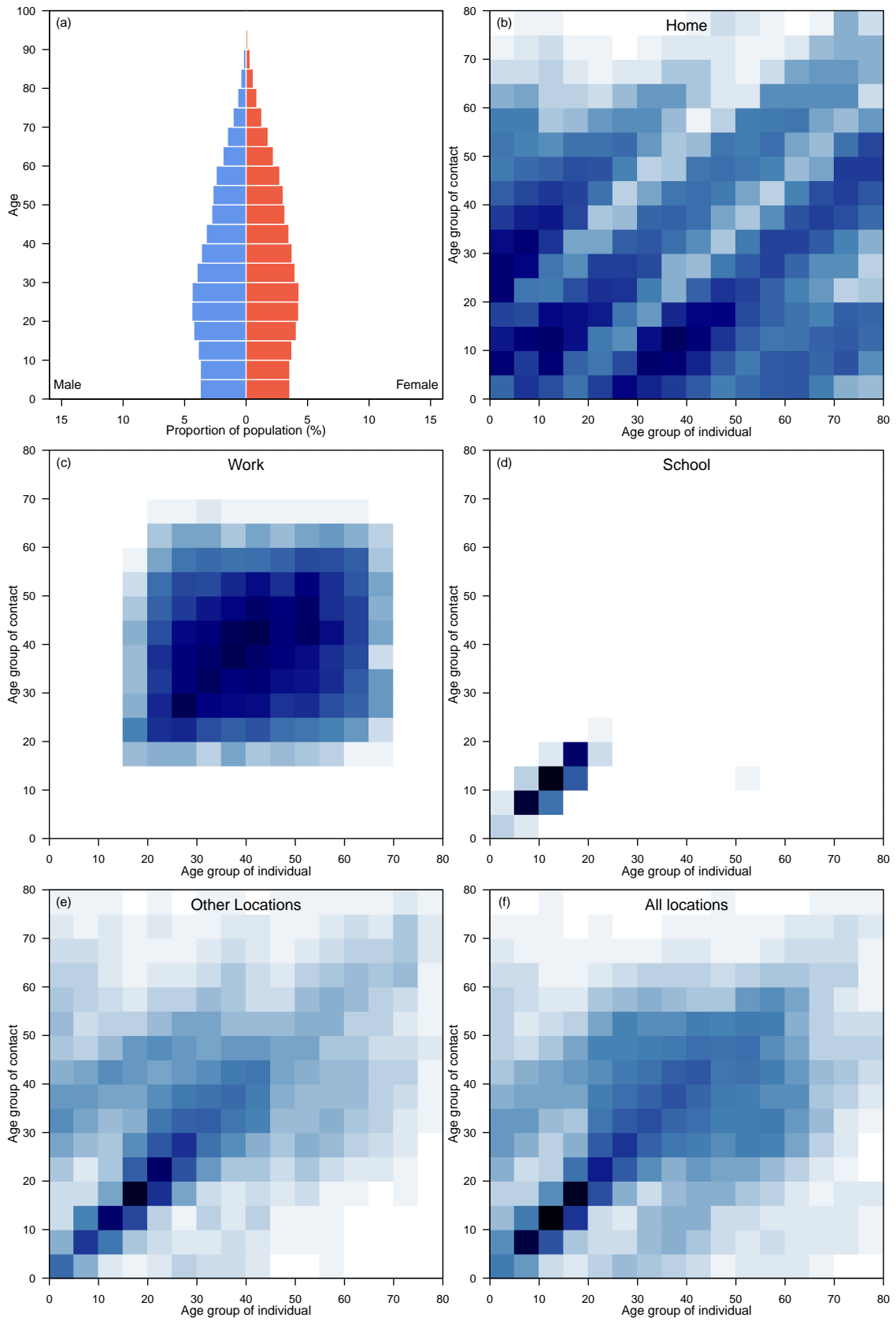
Comoros
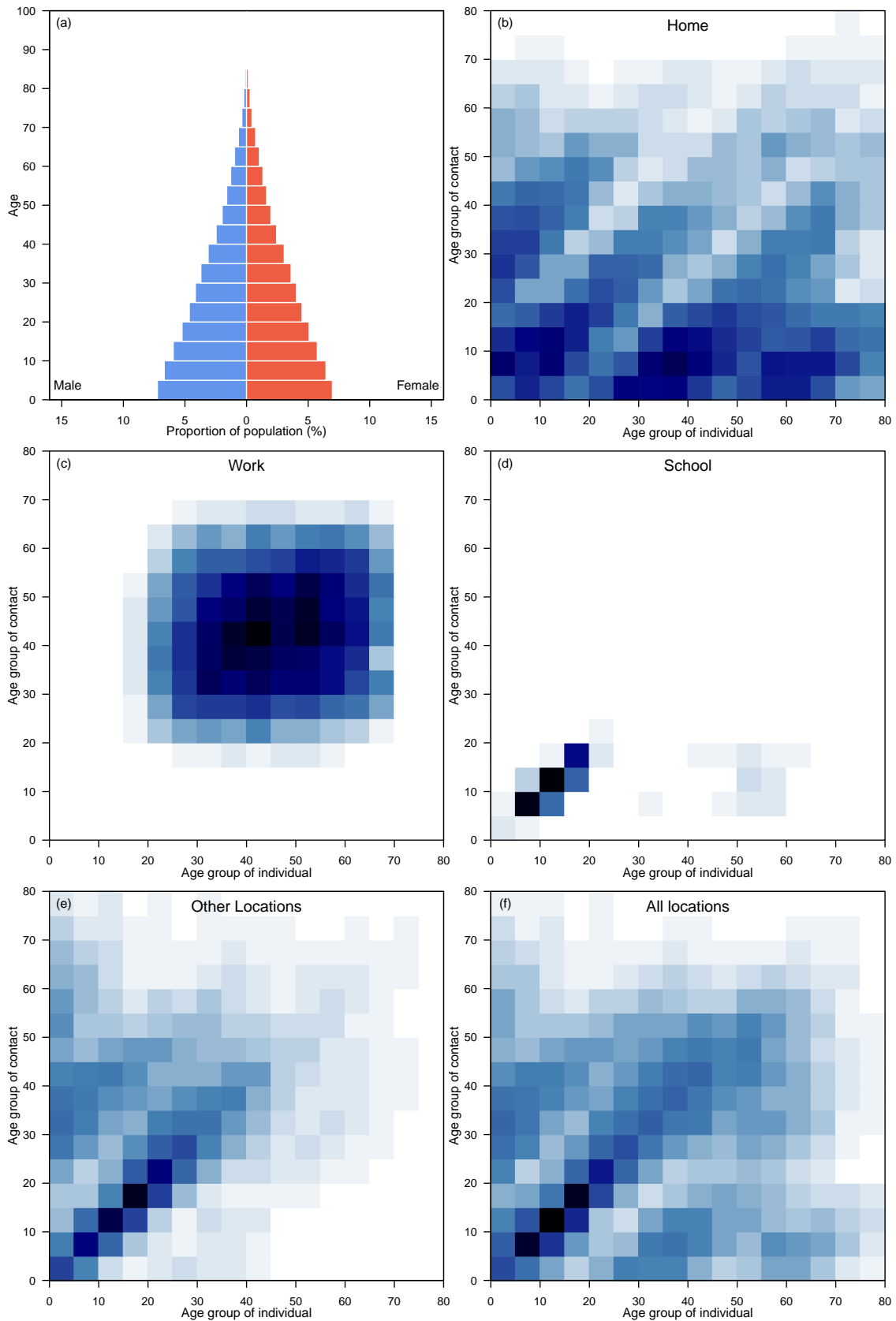
Congo
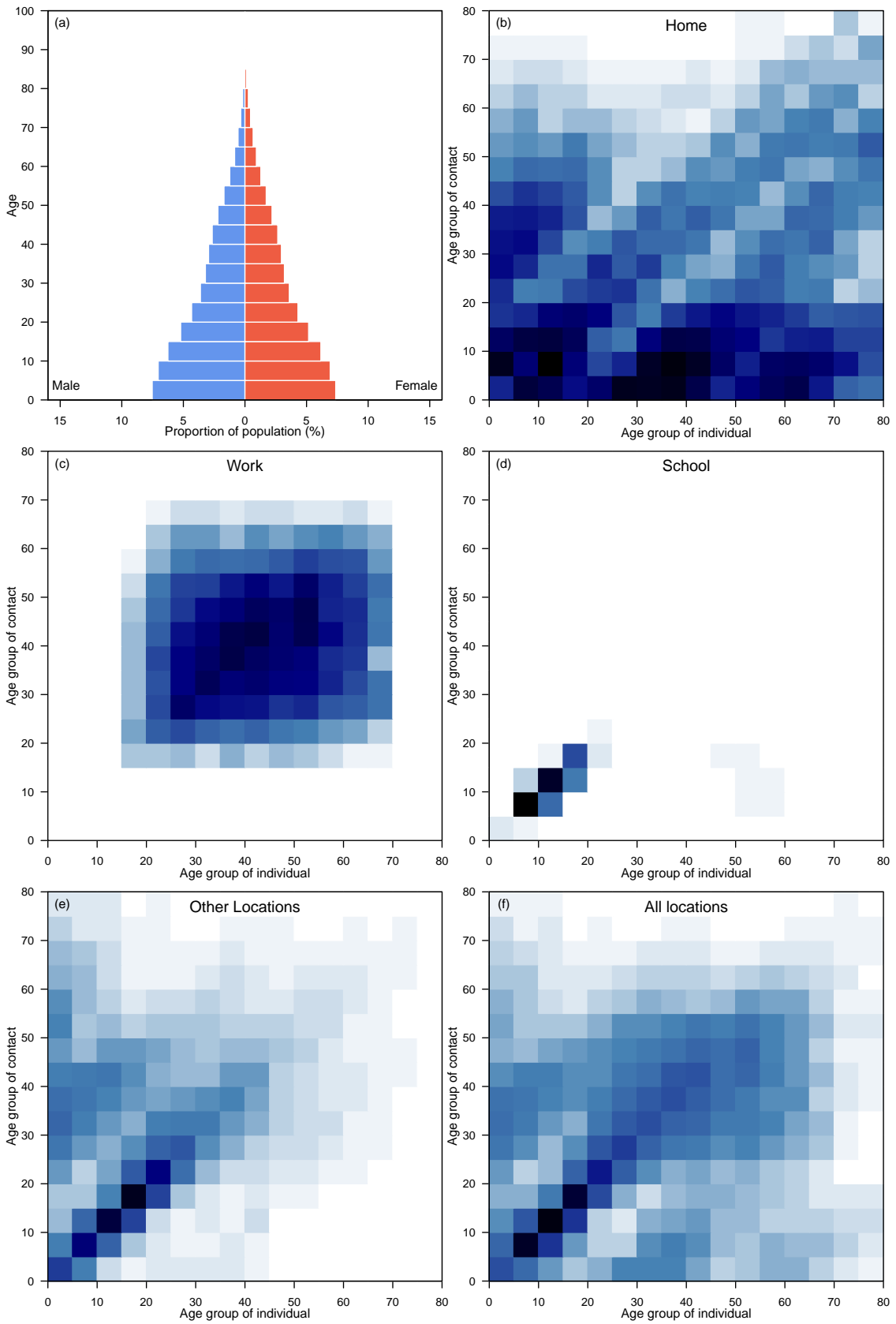
Costa Rica
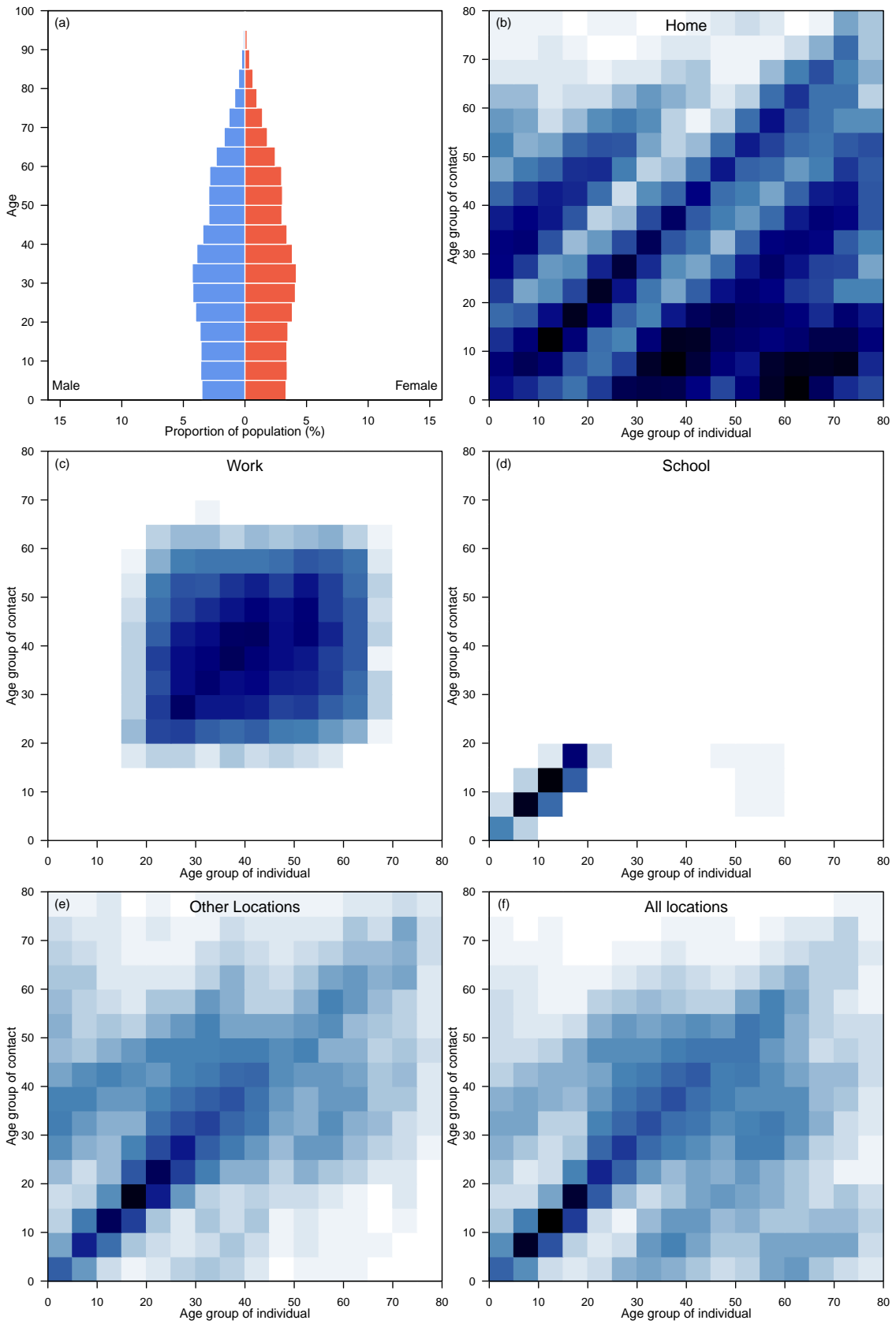
Côte d'Ivoire
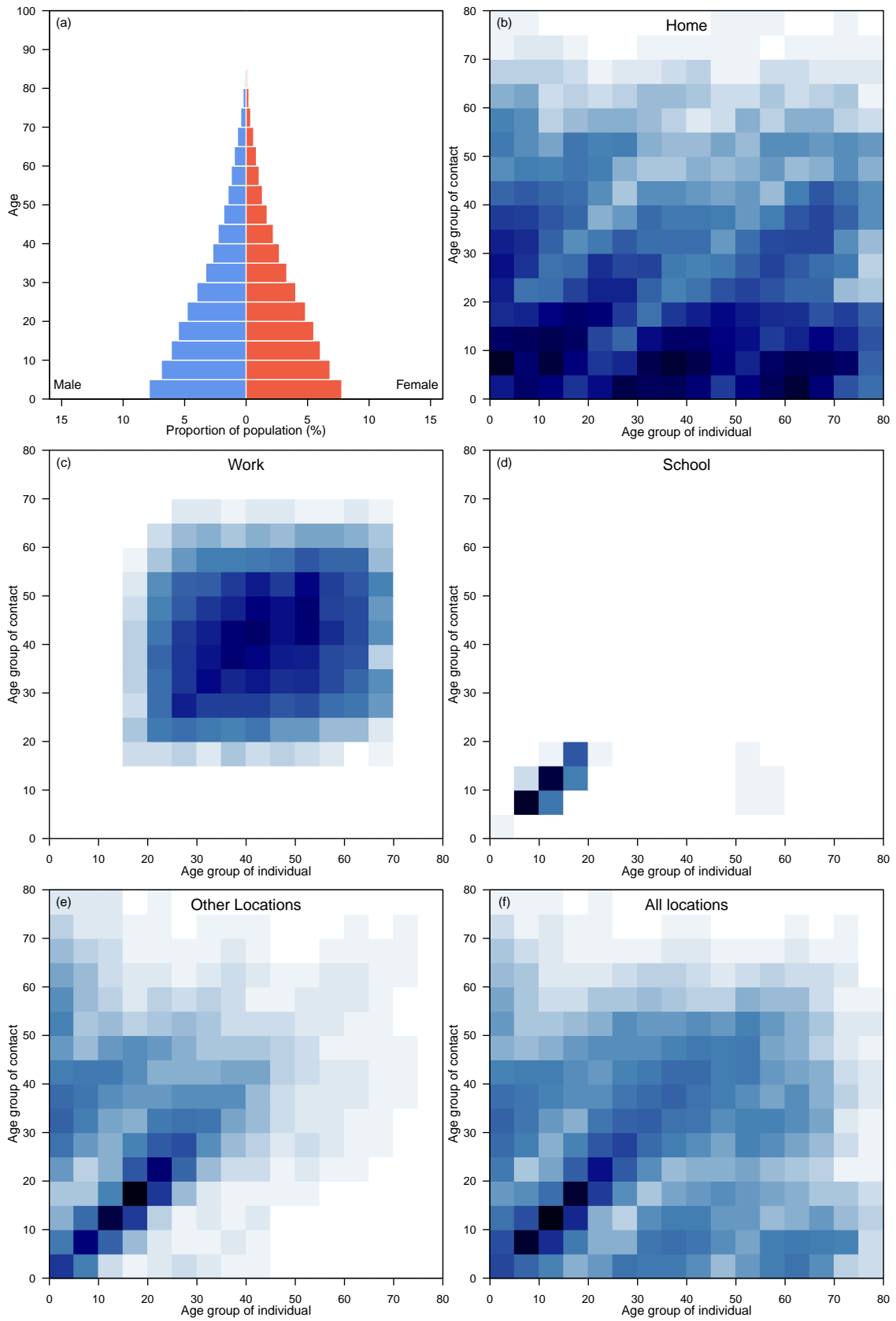
Croatia
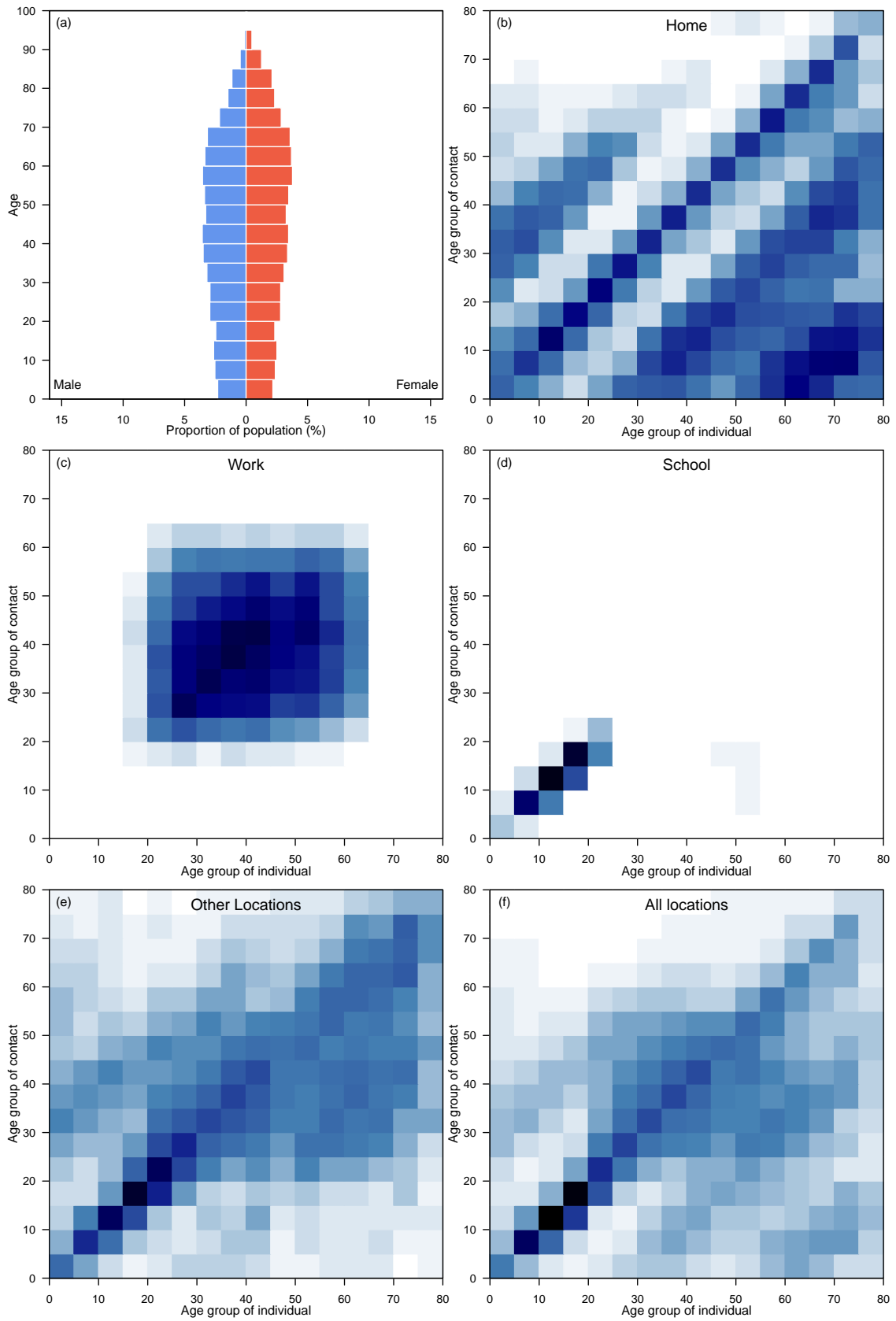
Cuba
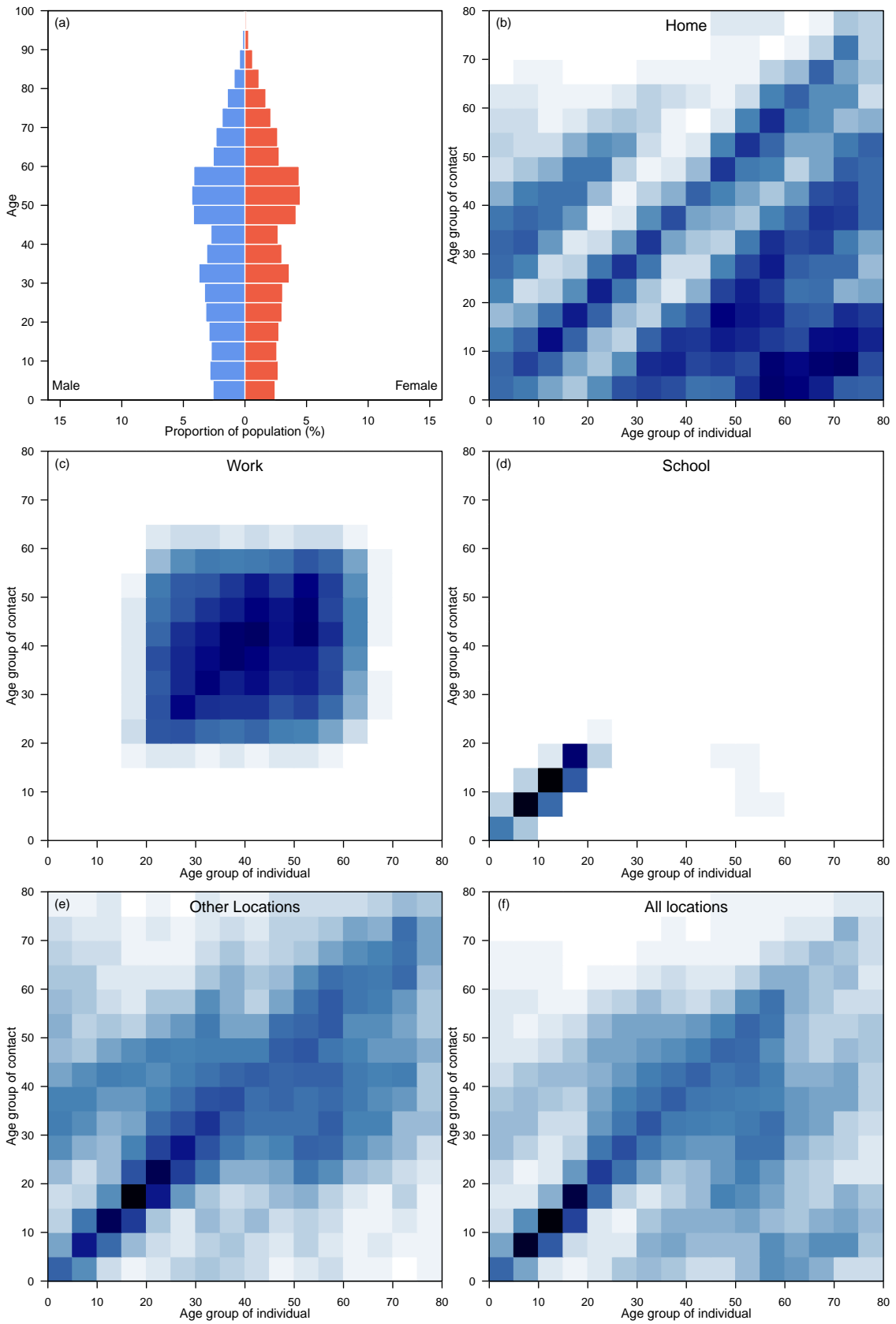
Cyprus
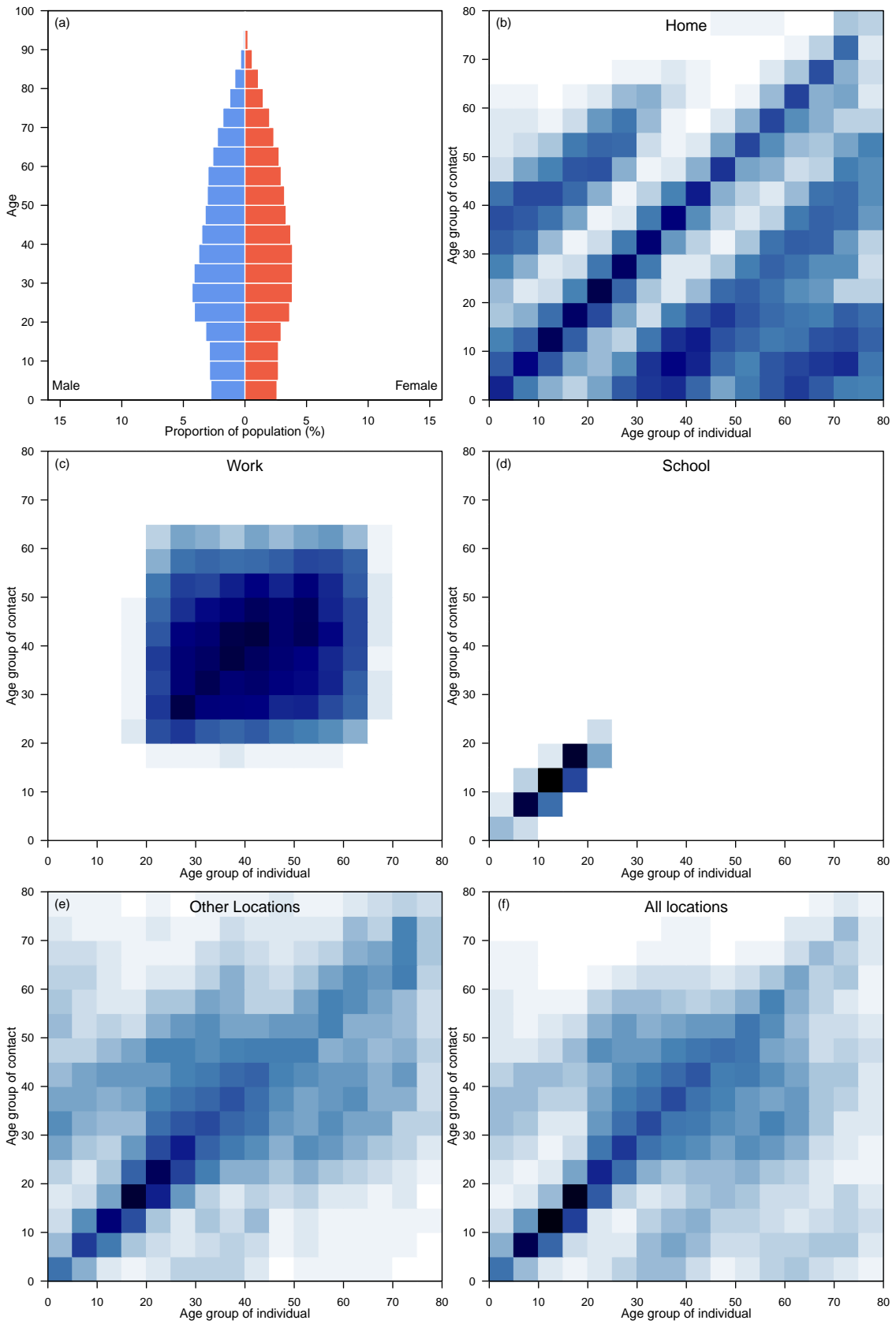
Czechia
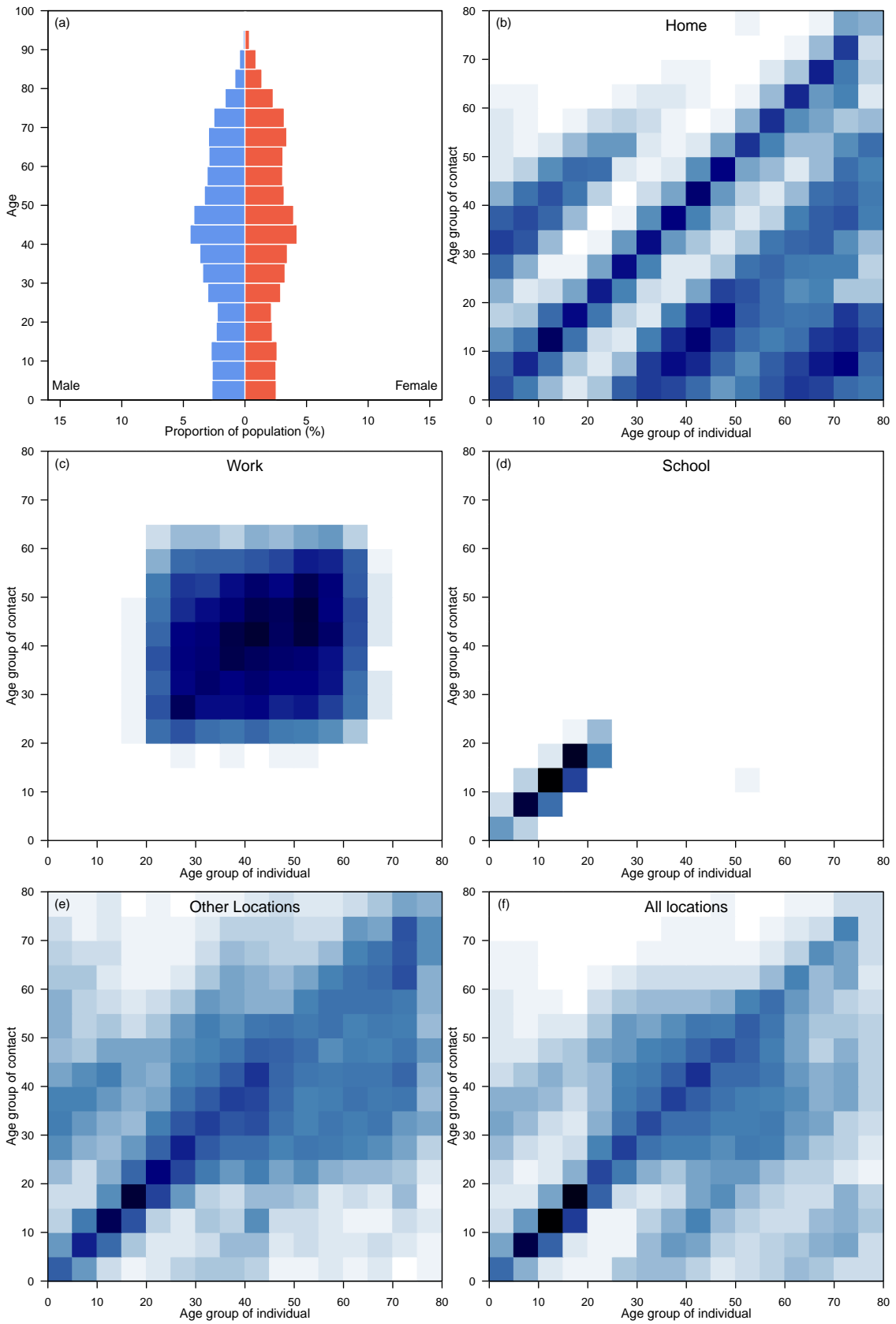
Dem. People's Republic of Korea
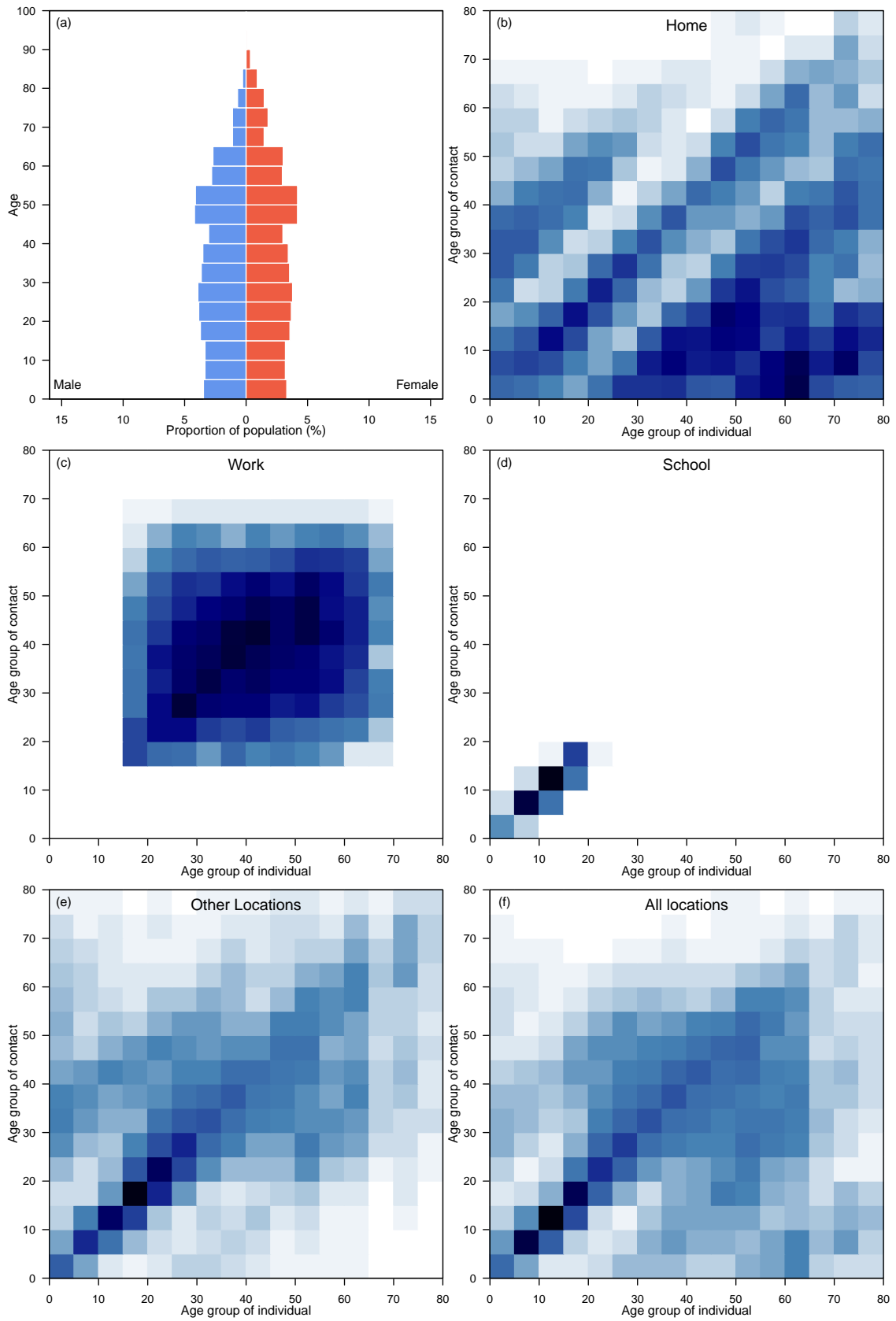
Democratic Republic of the Congo
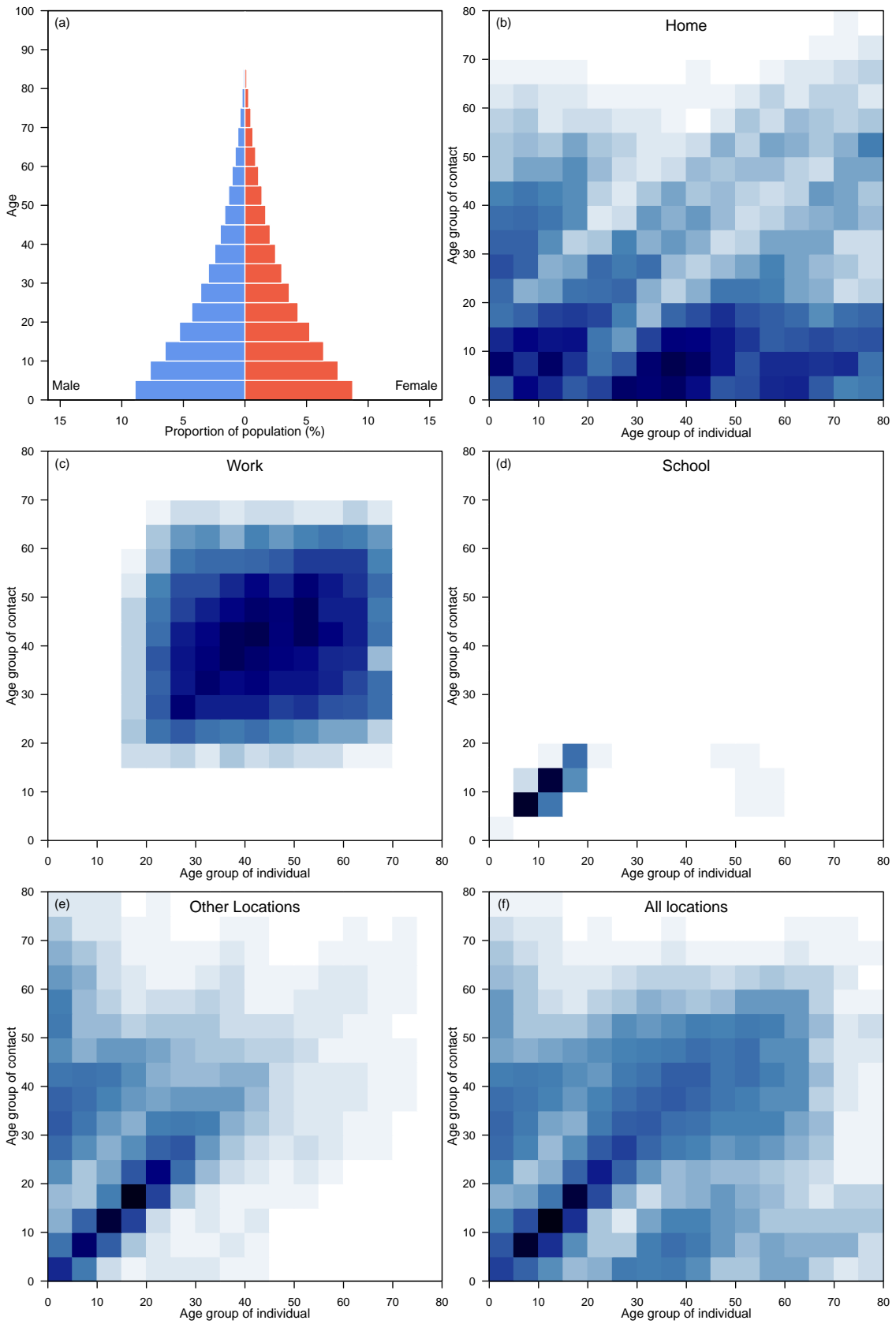
Denmark
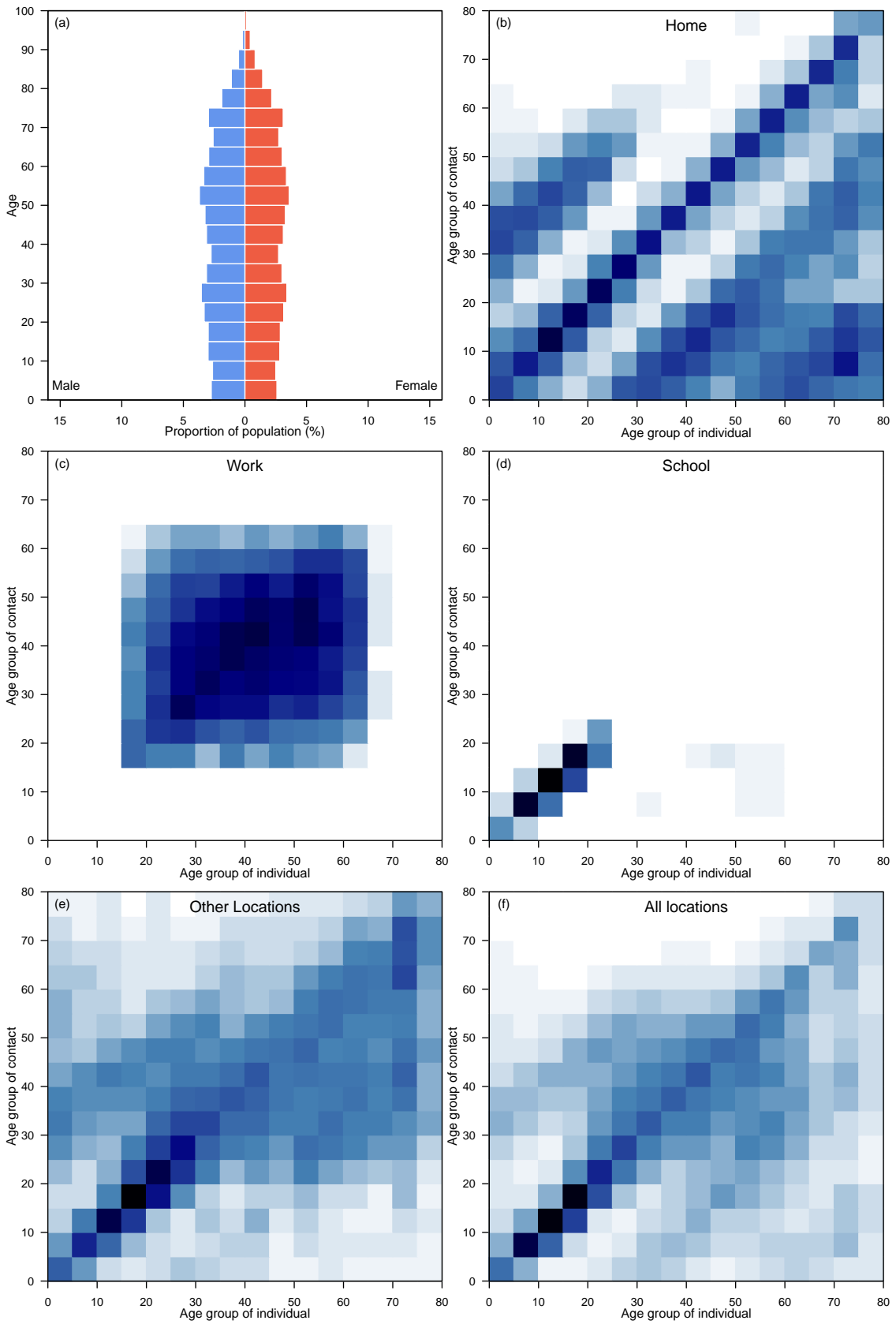
Djibouti
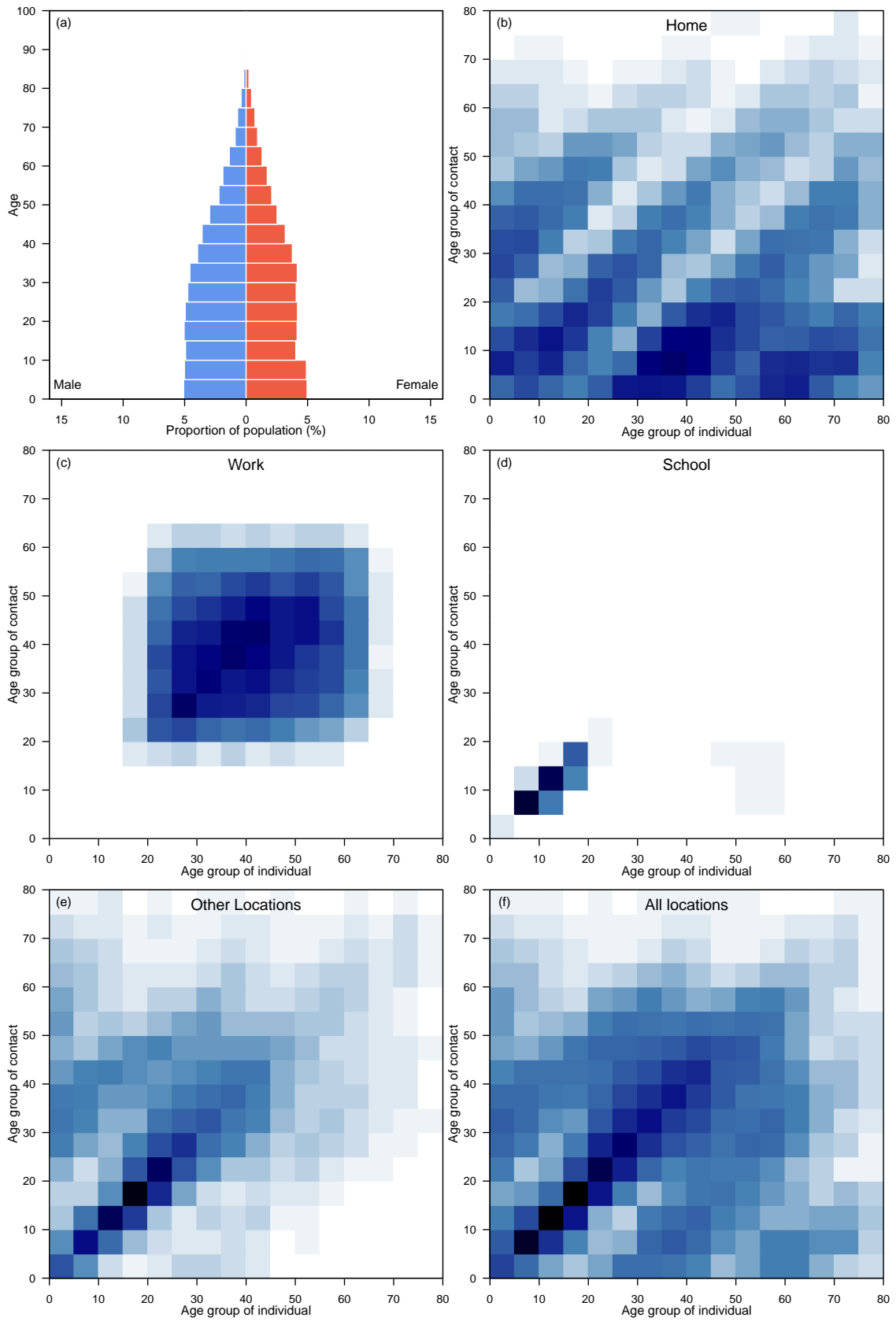
Dominican Republic
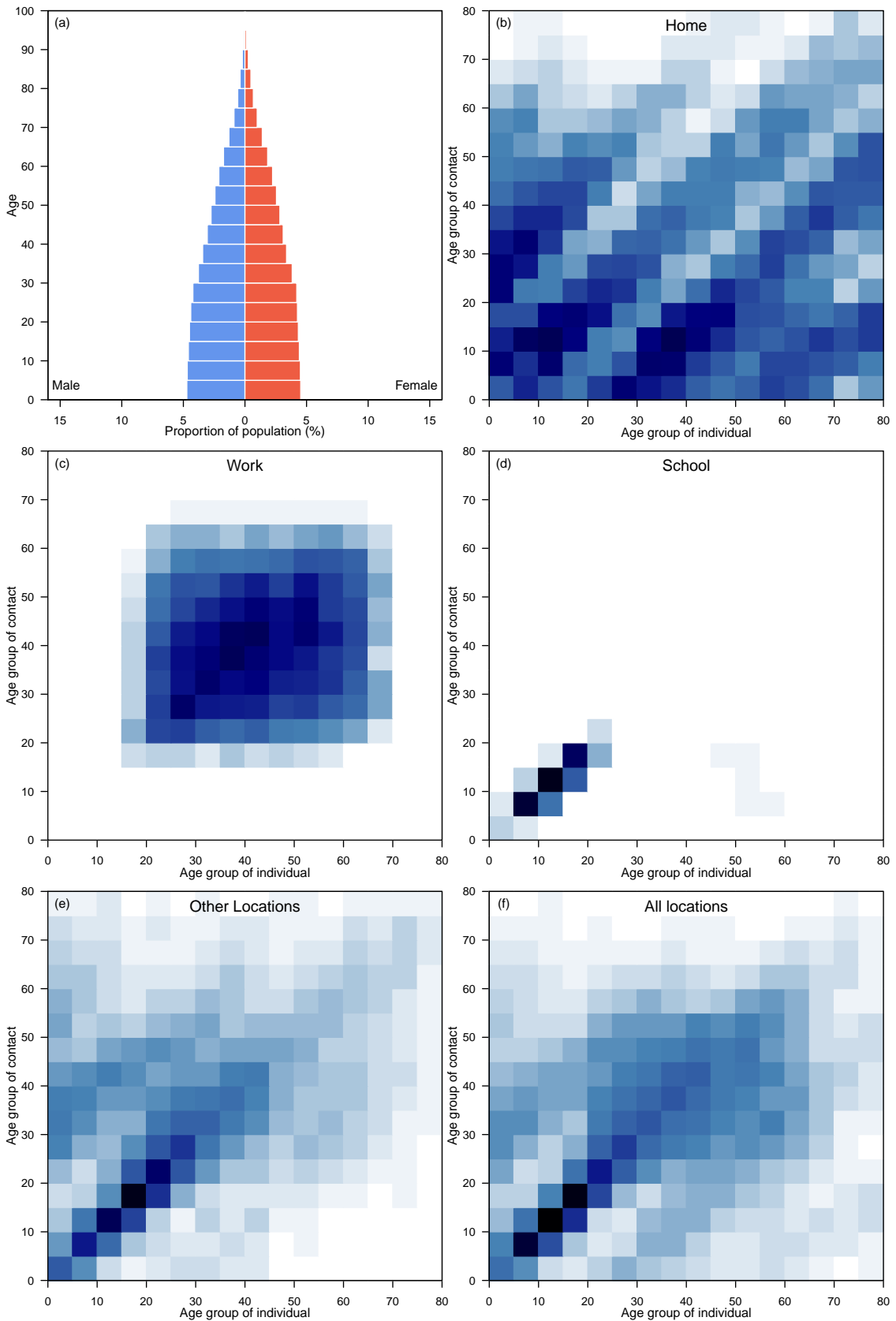
Ecuador
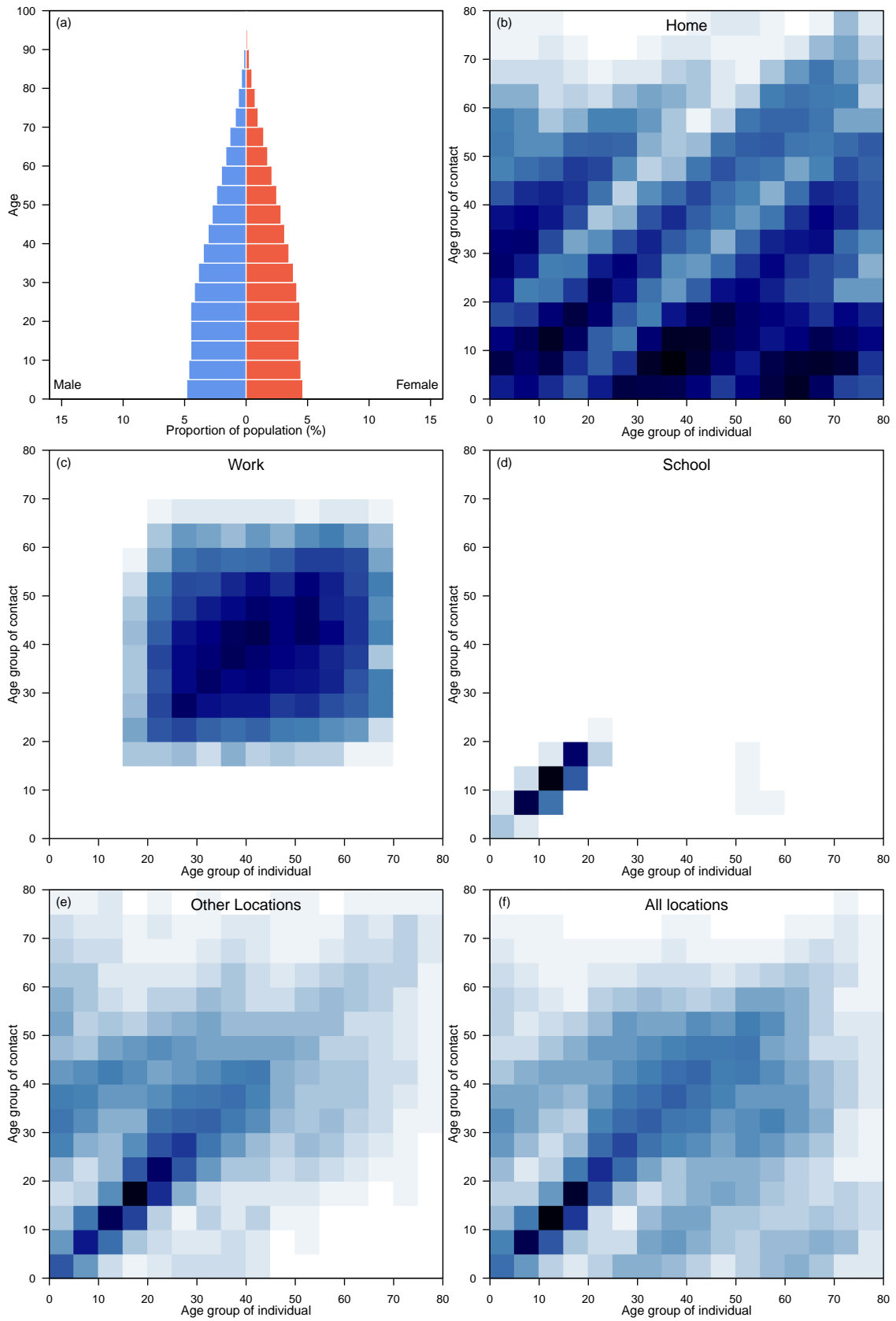
Egypt
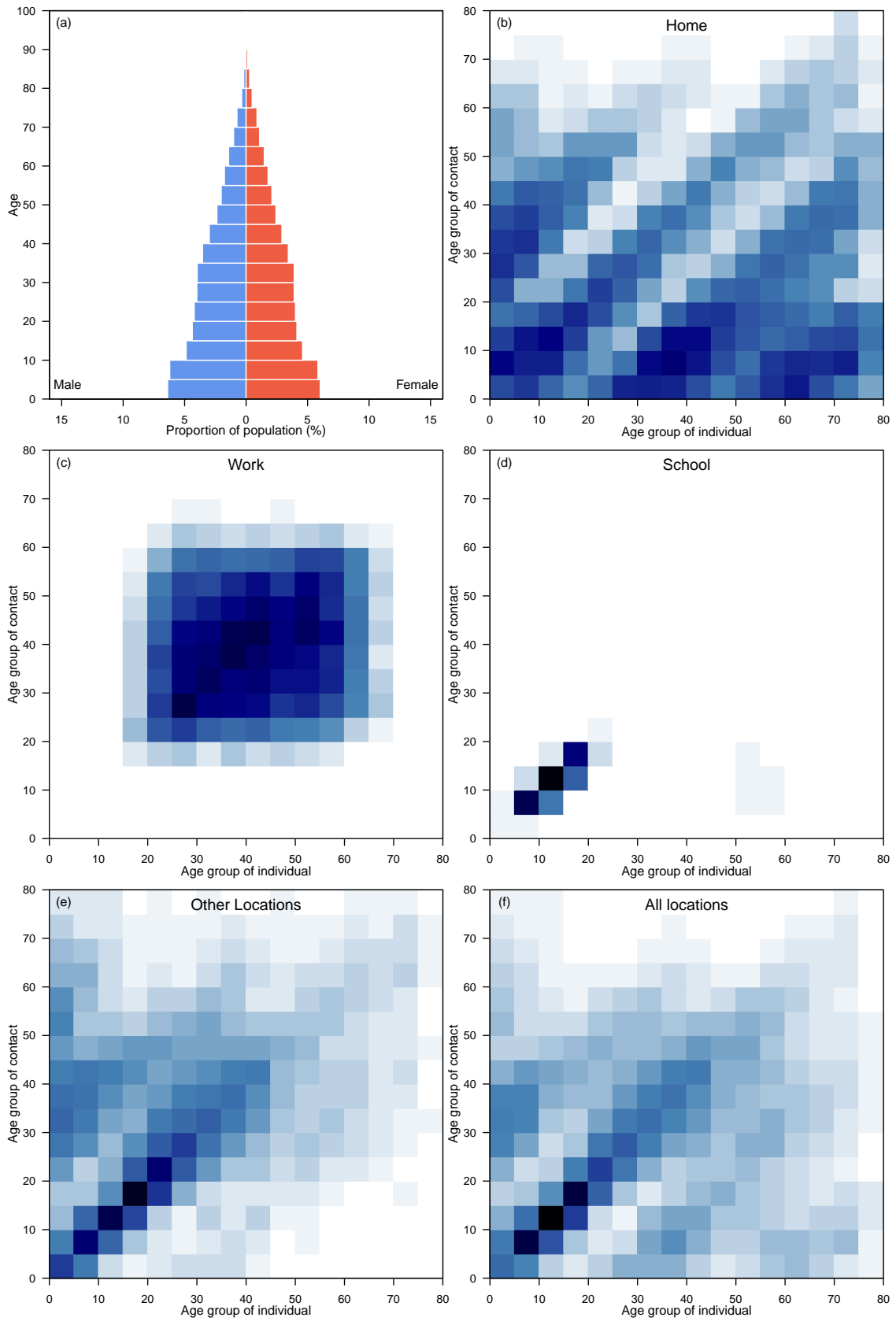
El Salvador
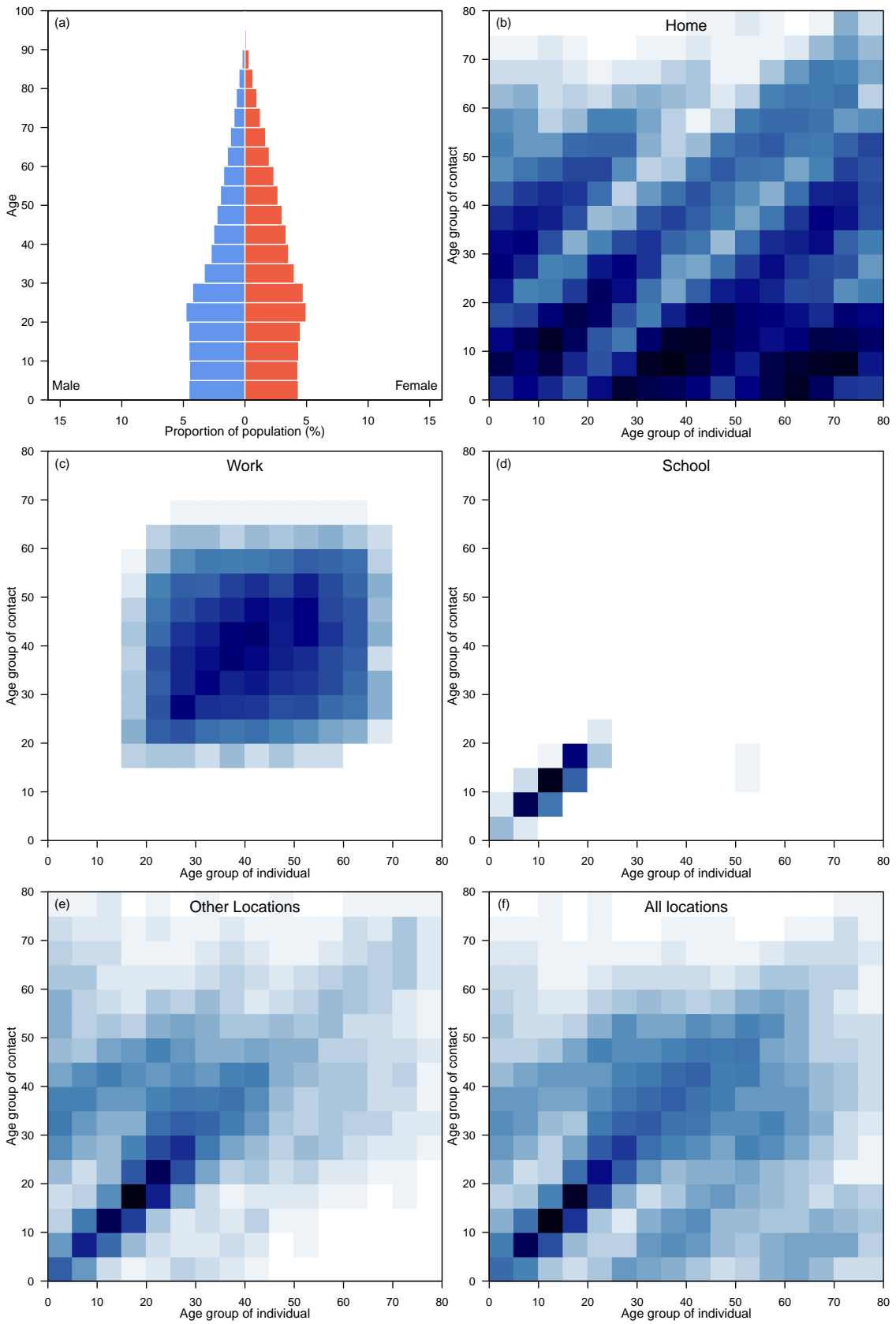
Equatorial Guinea
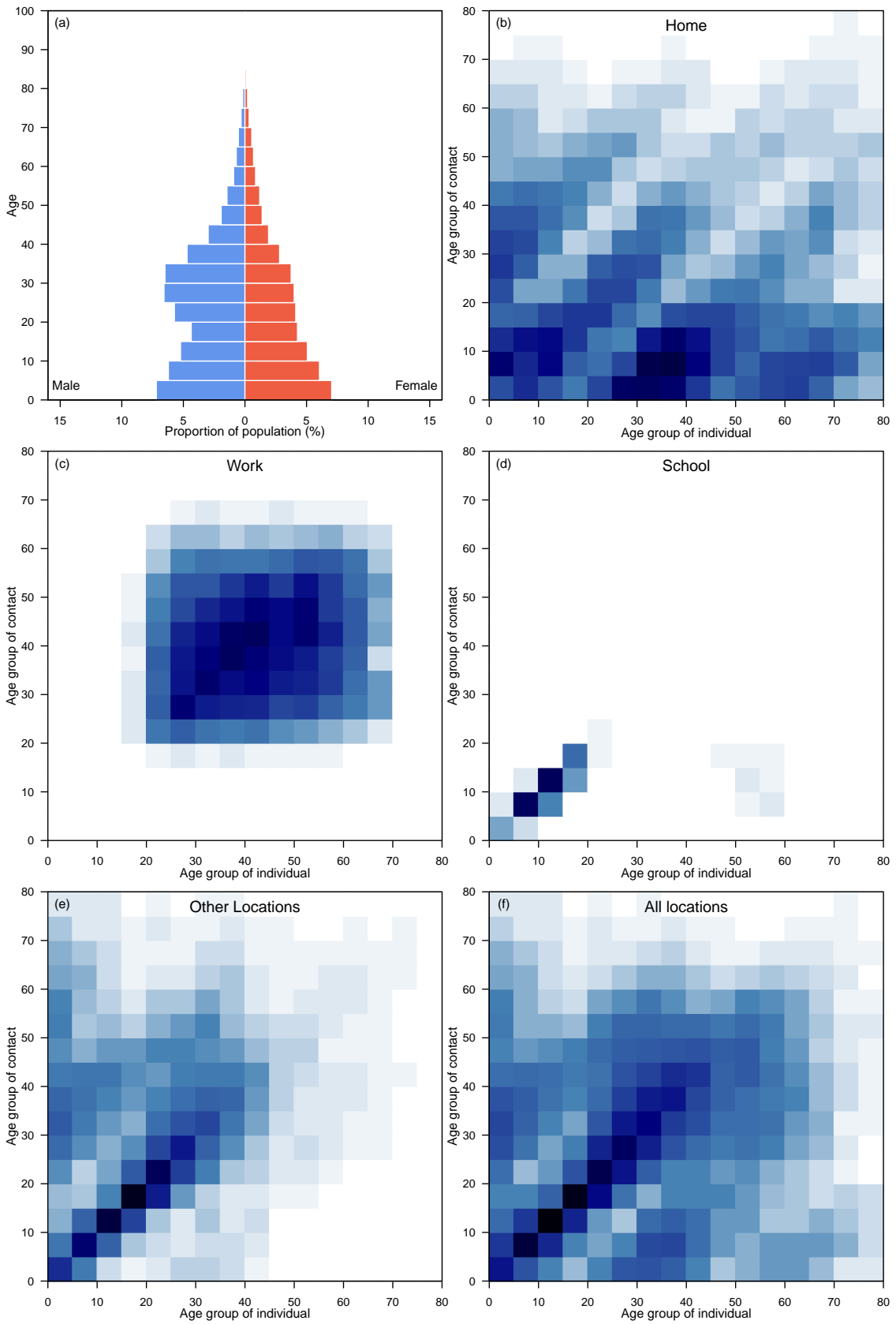
Eritrea
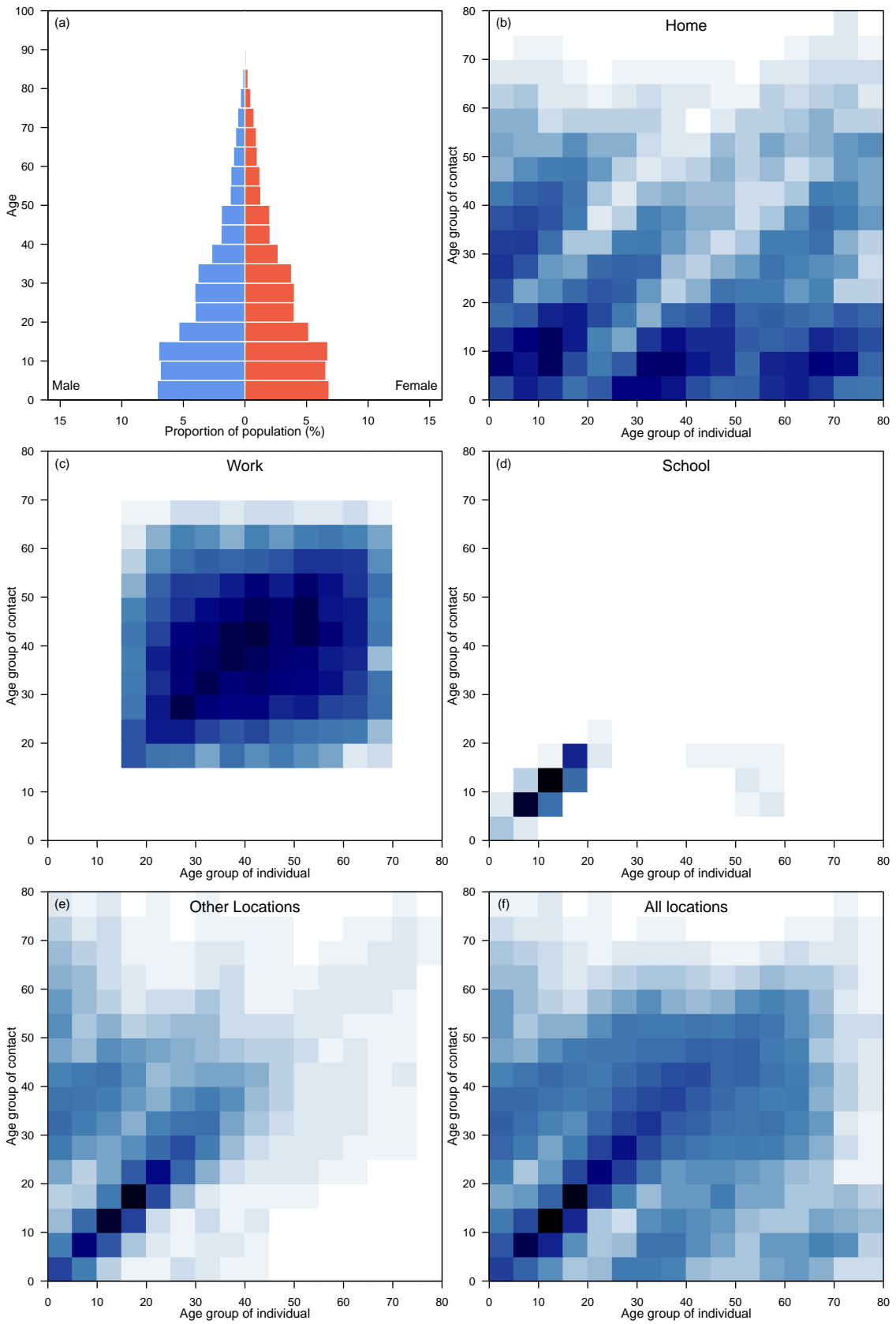
Estonia

Eswatini

Ethiopia

Fiji

(a)

(b) Home

(c) Work

(d) School

(e) Other Locations

(f) All locations
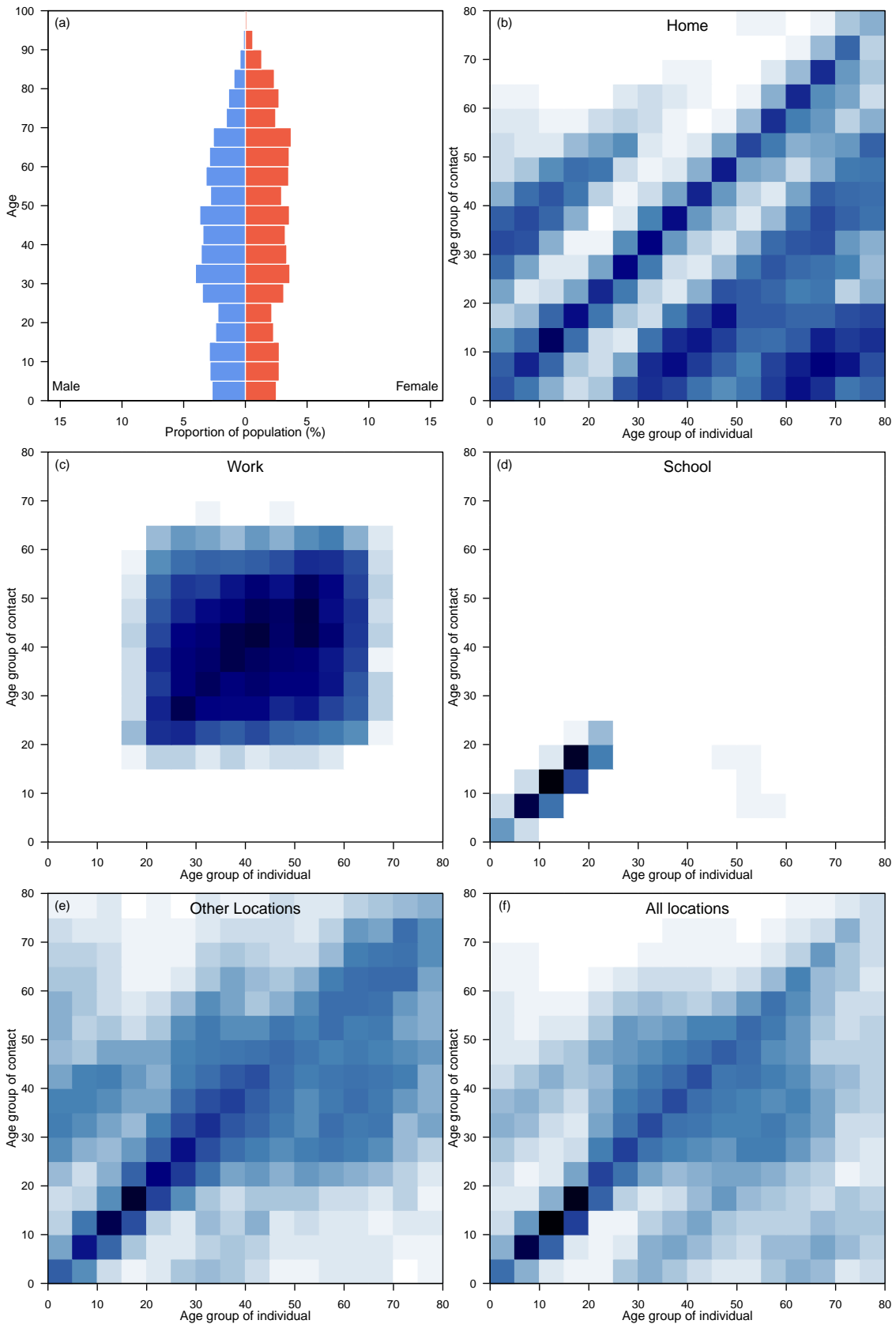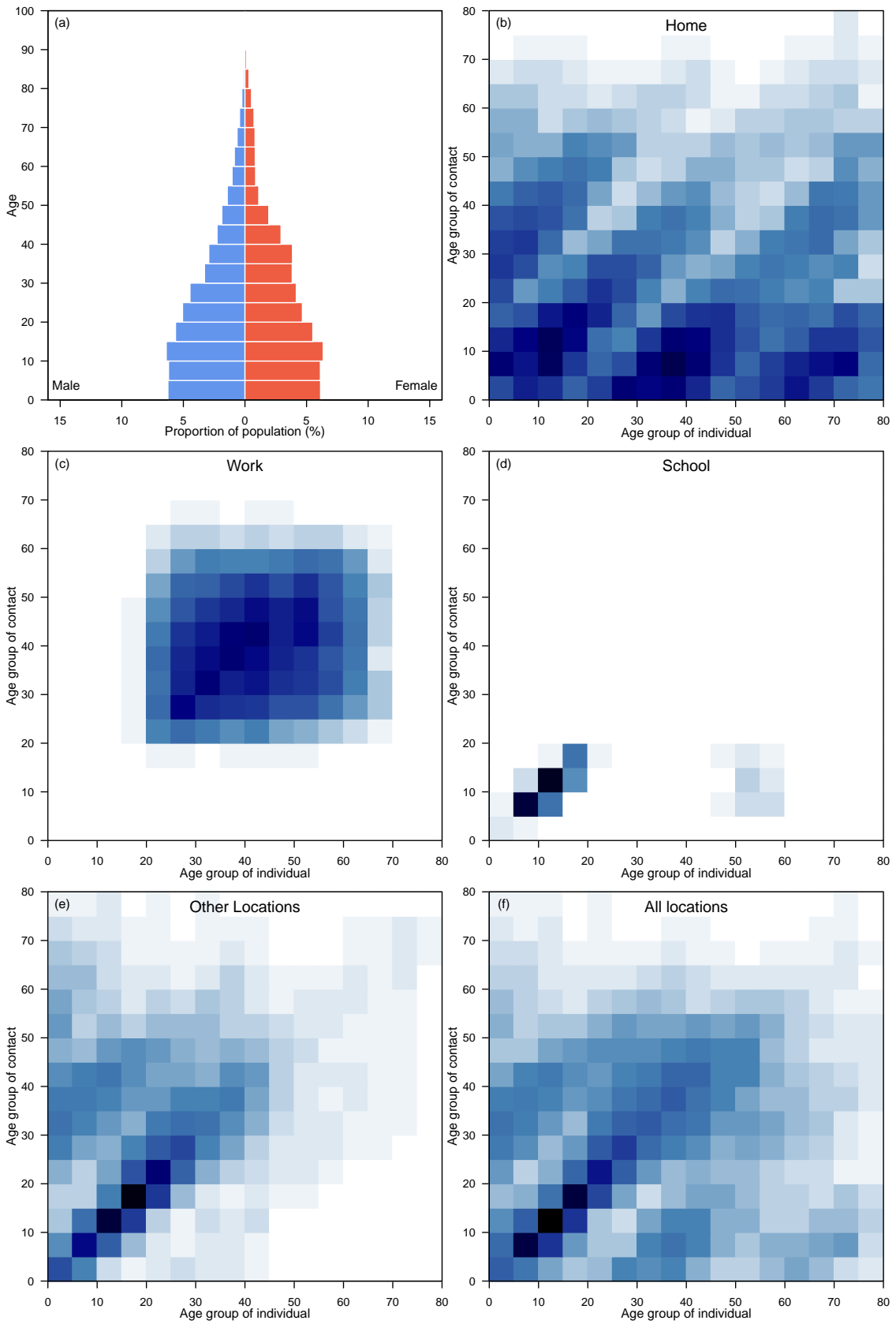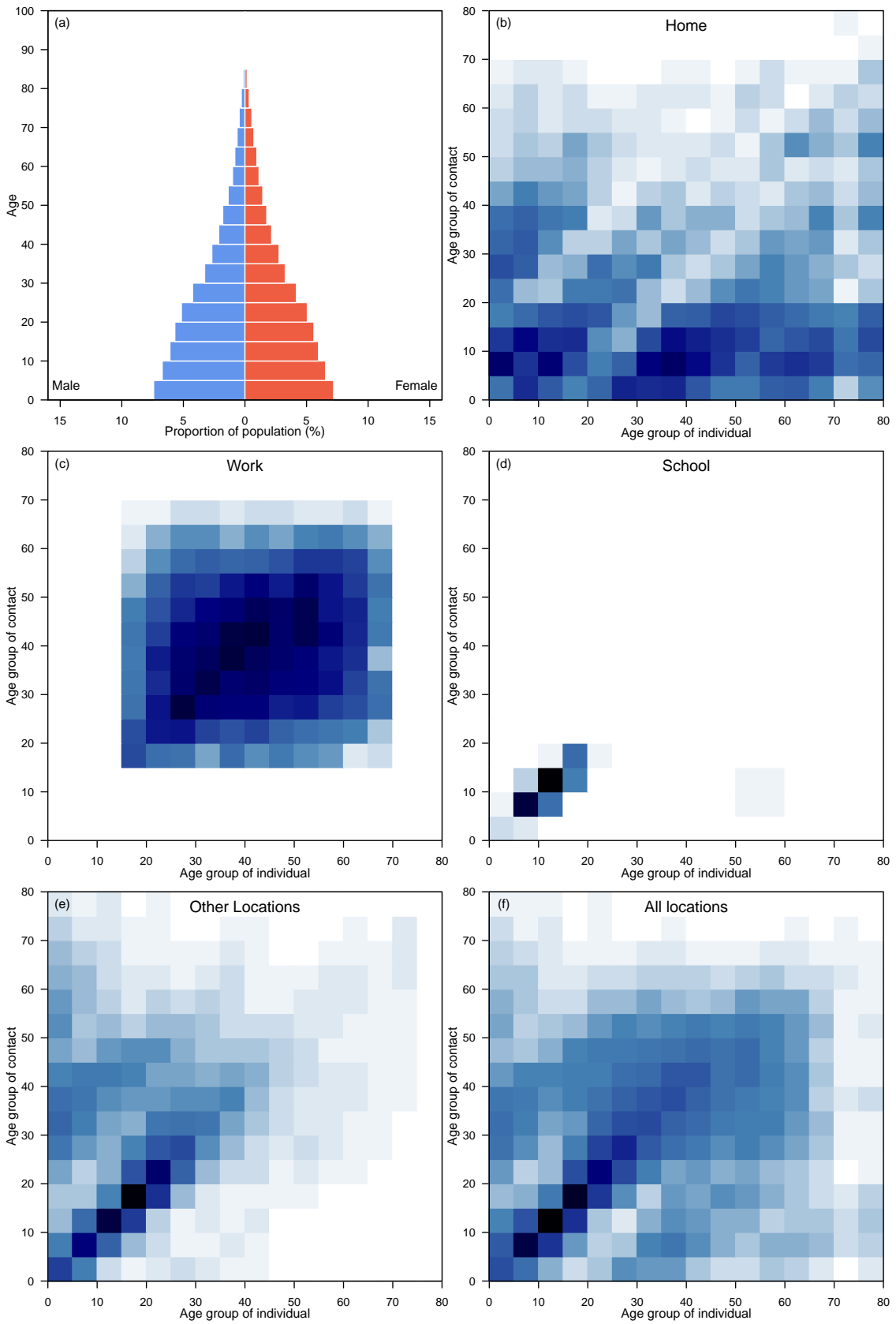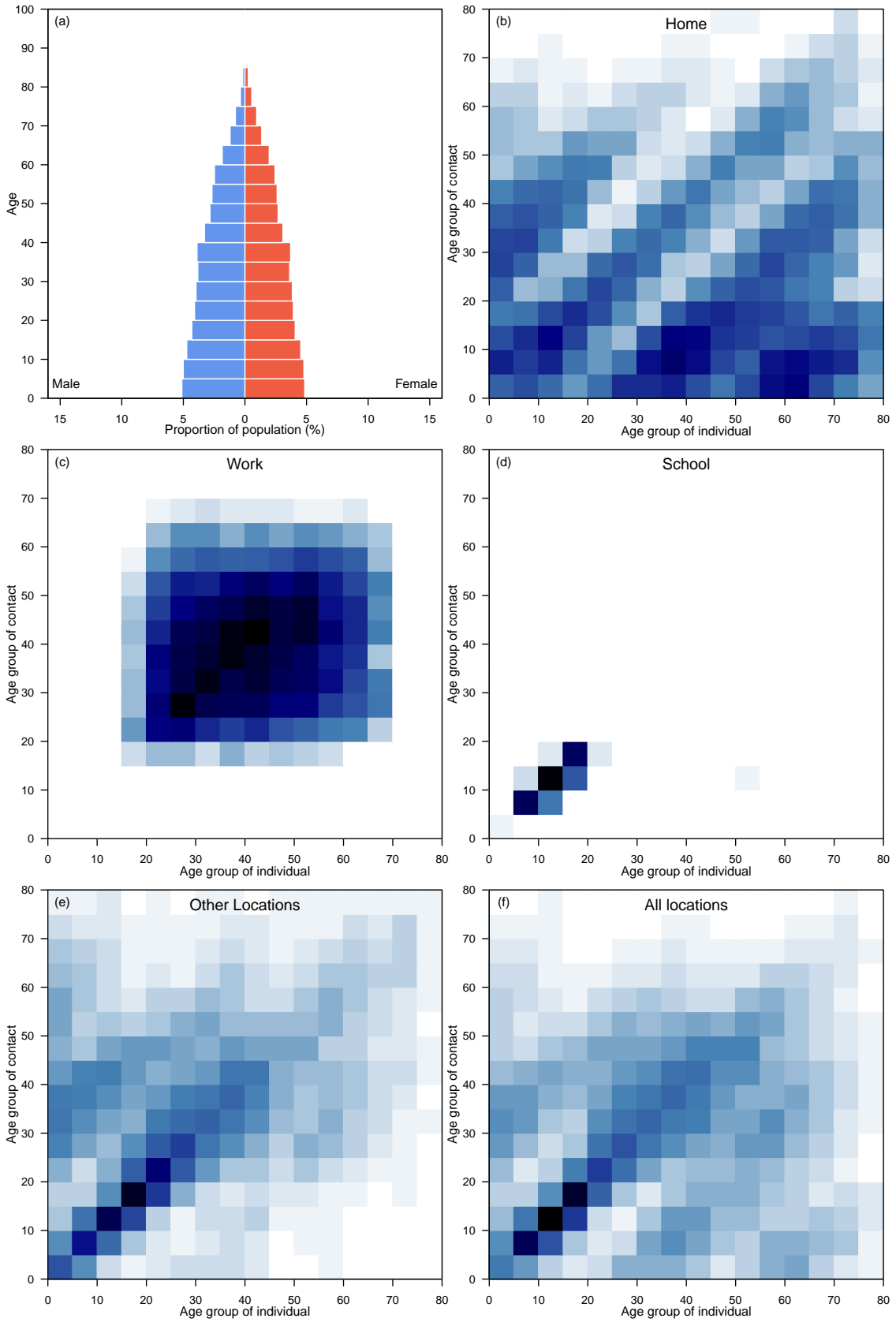
Finland
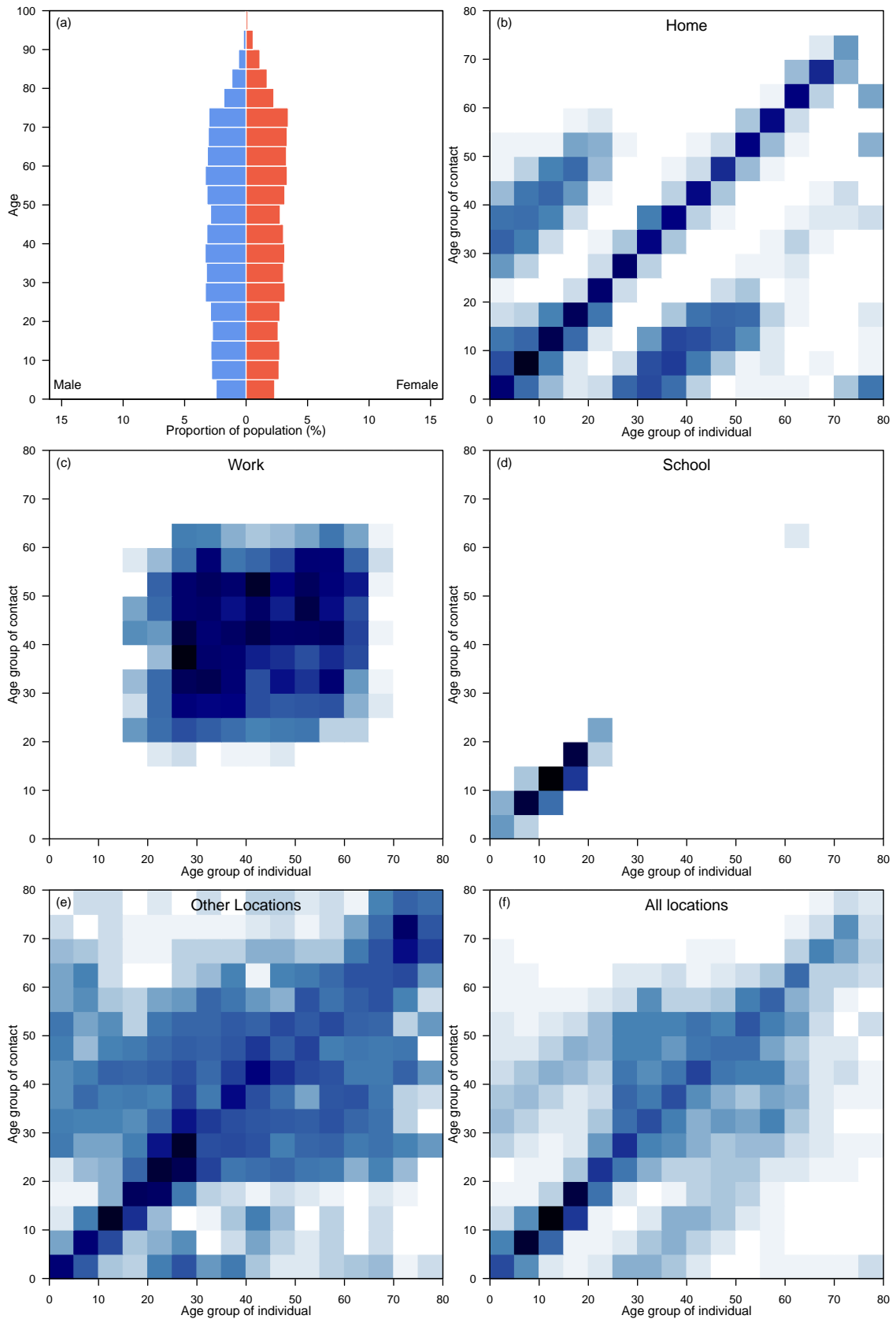
France
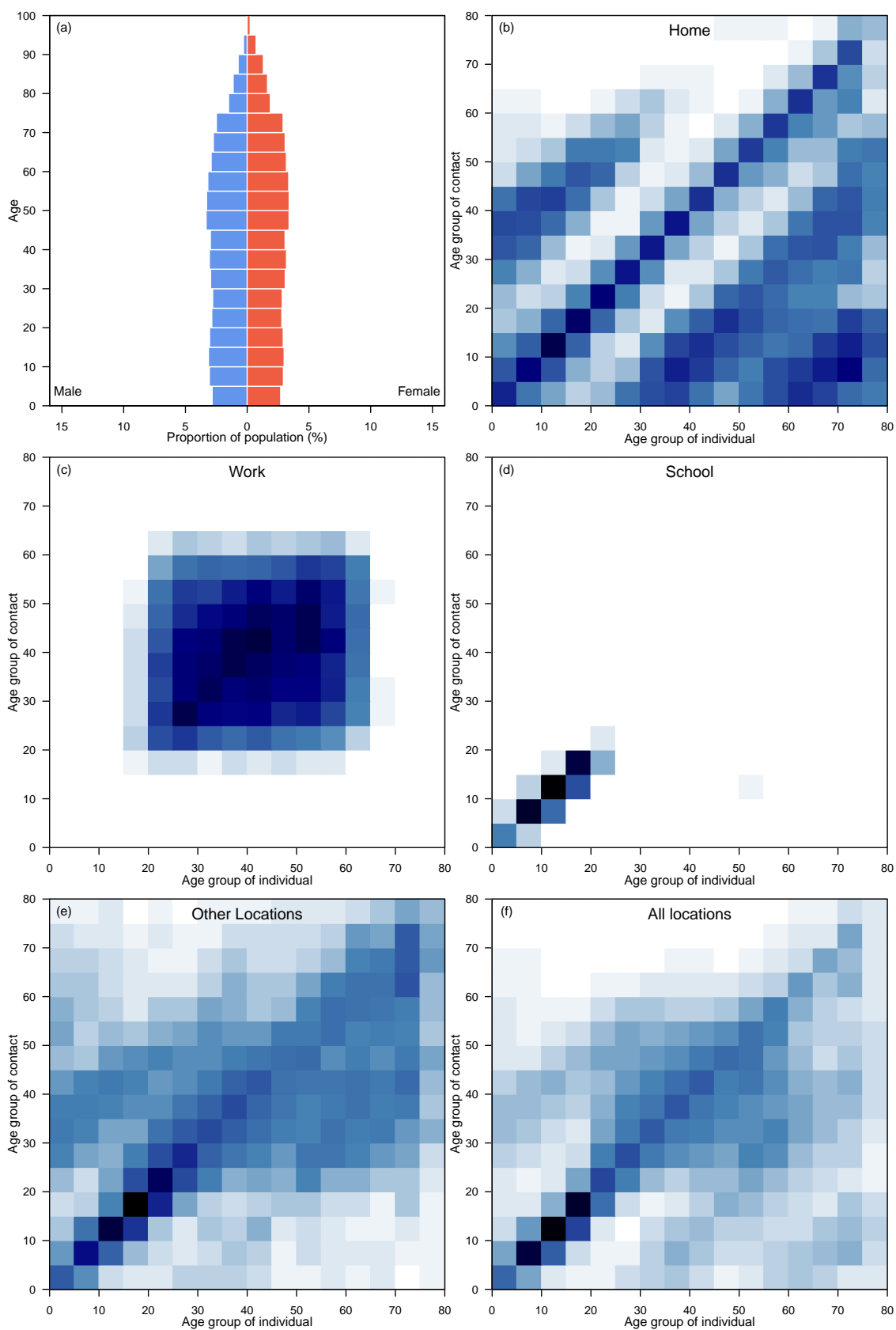
Gabon
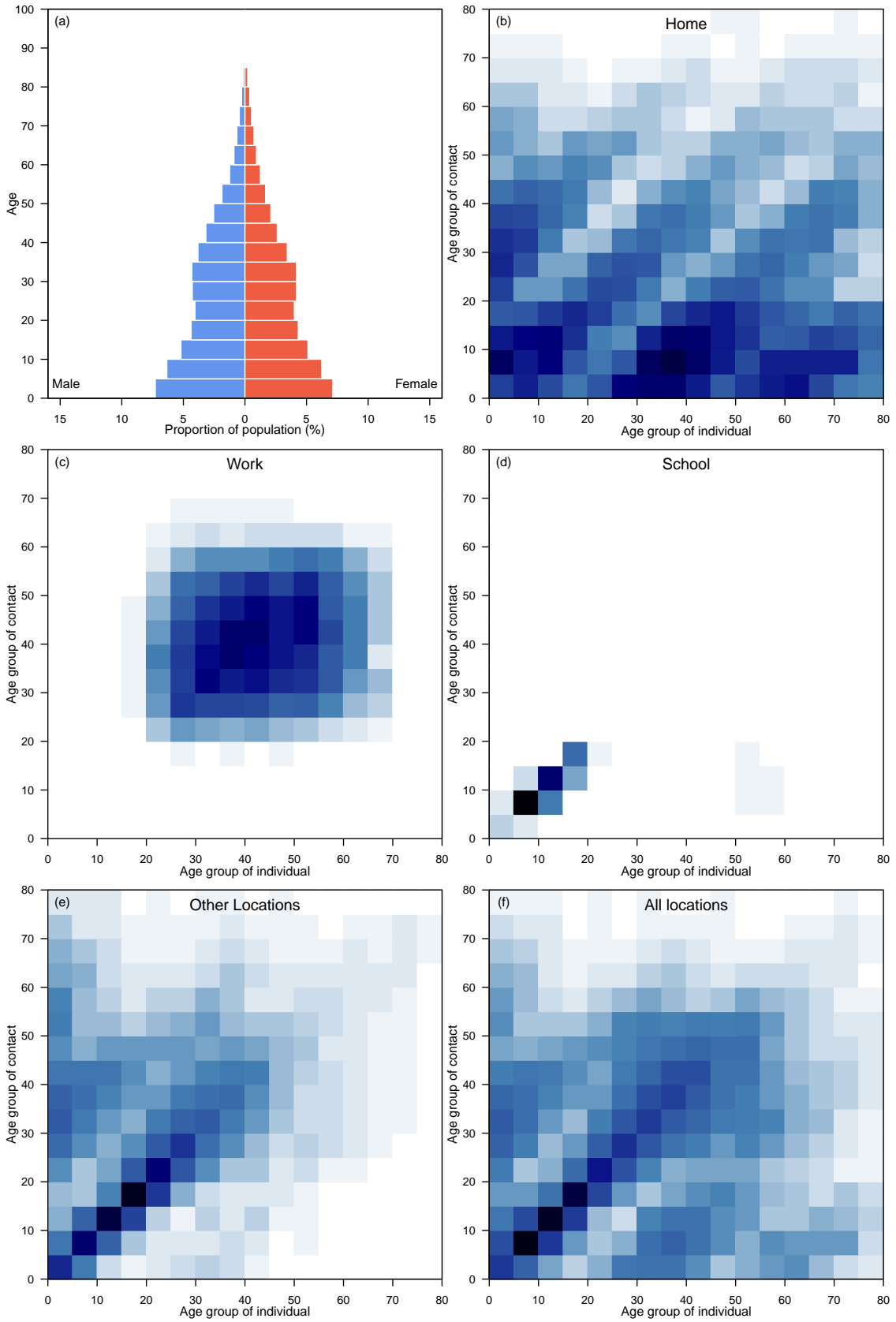
Gambia
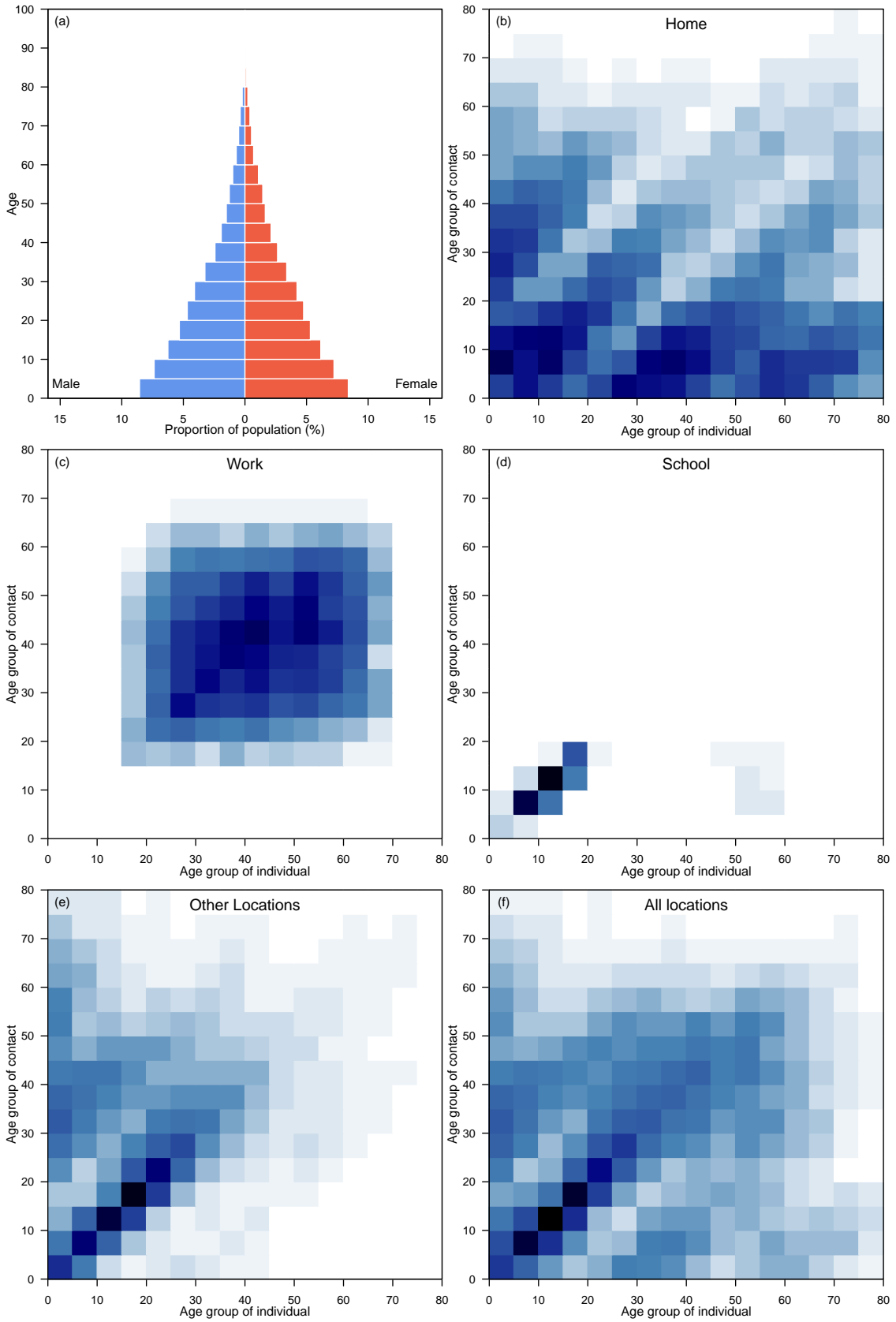
Georgia
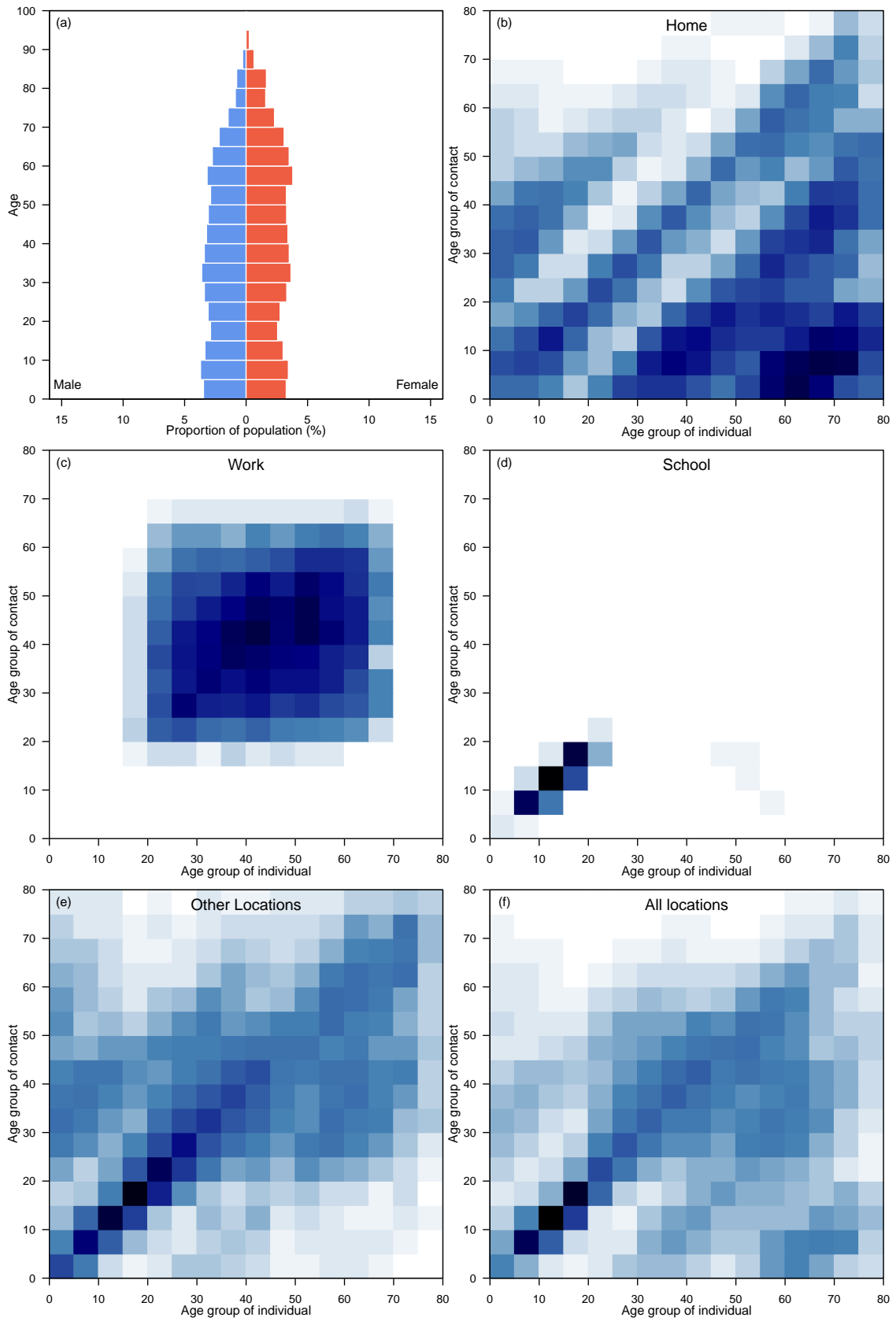
Germany
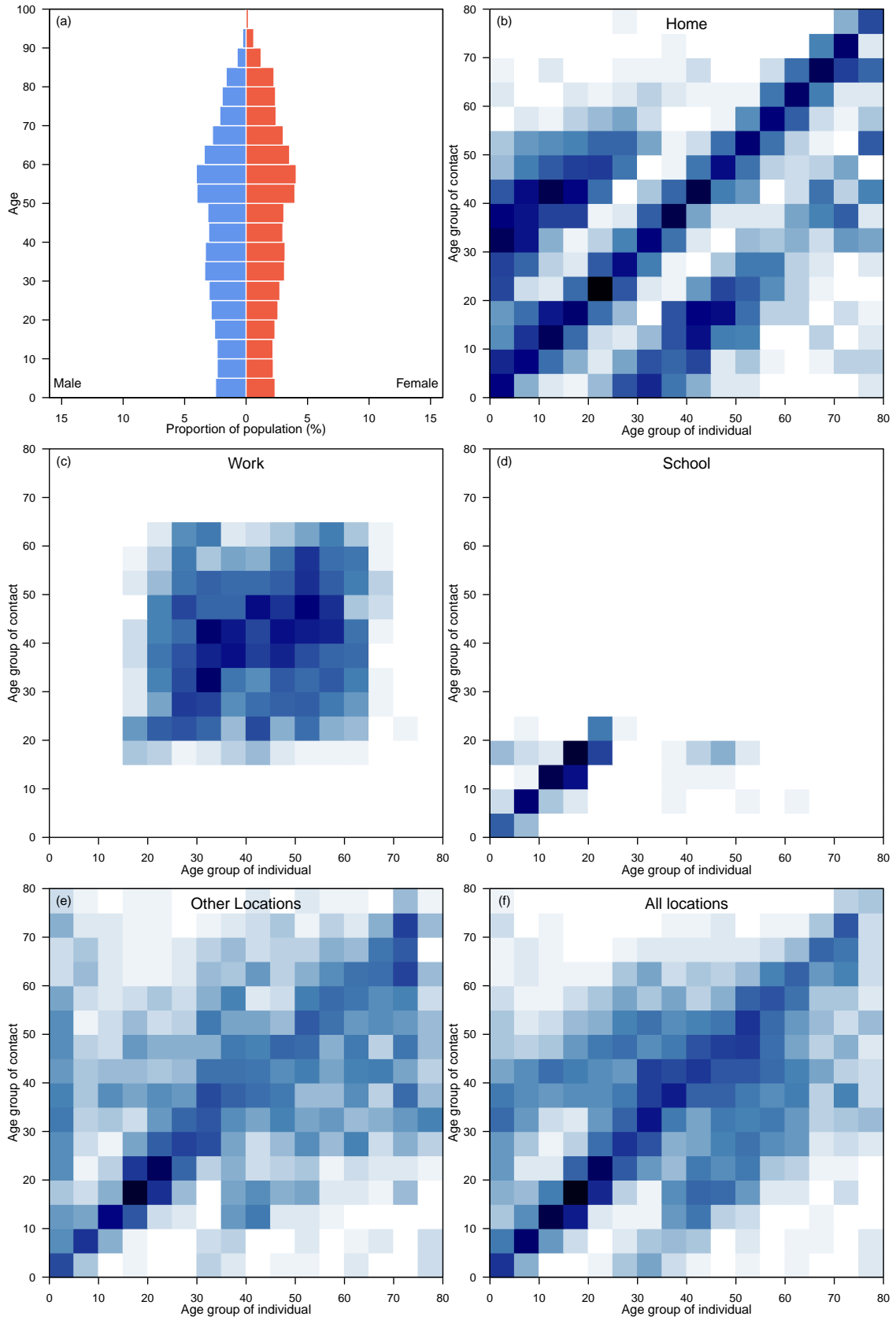
Ghana
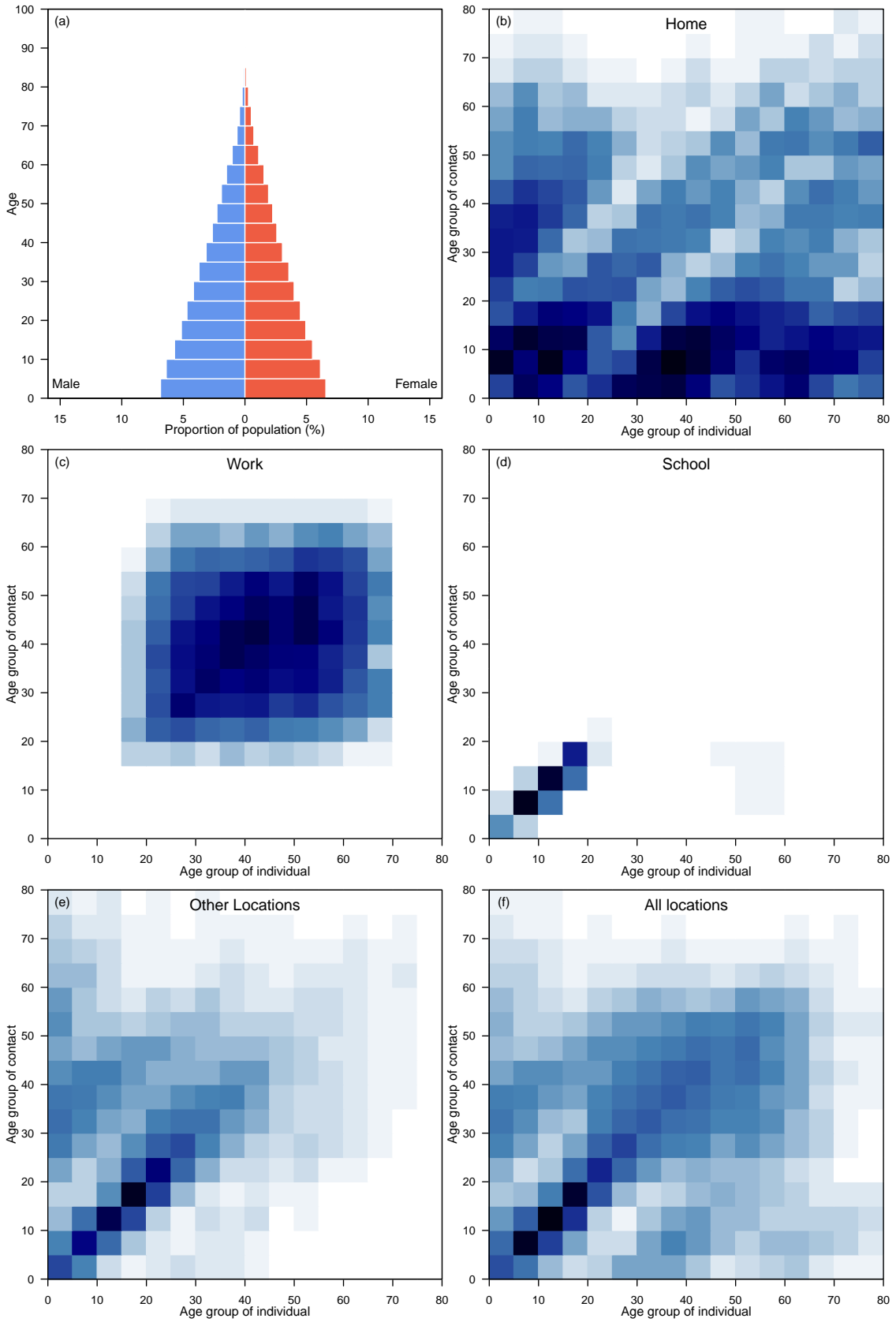
Greece
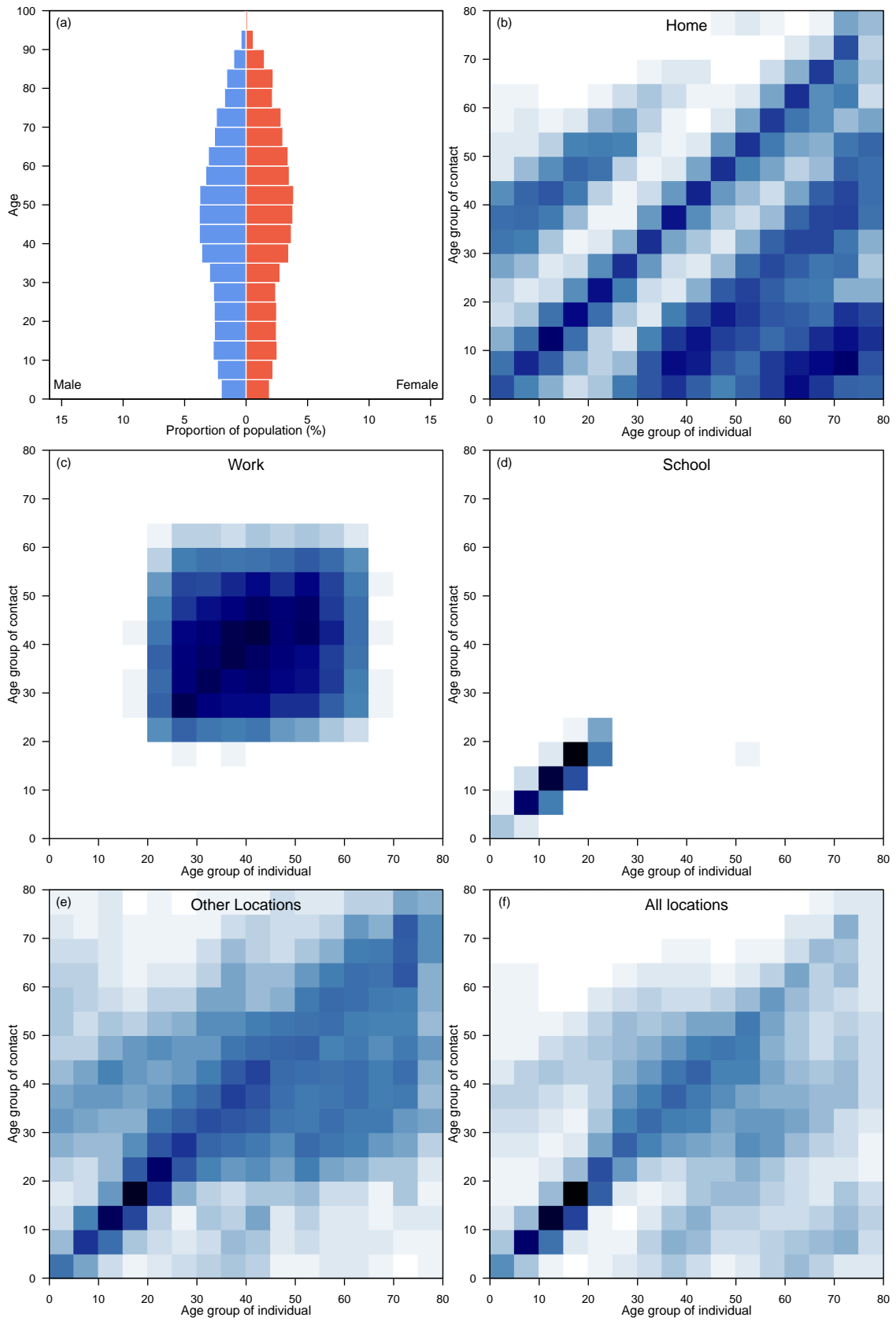
Guatemala
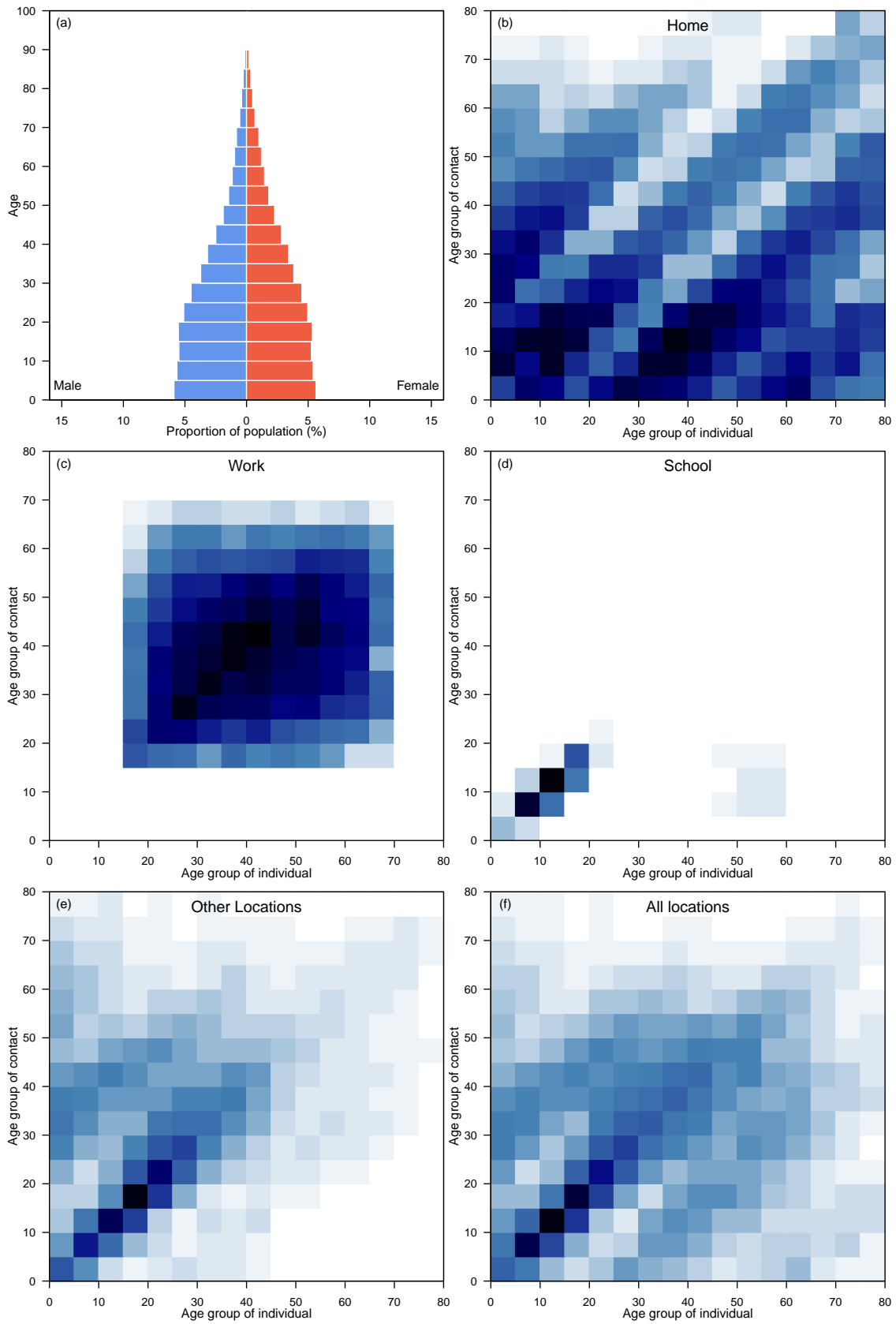
Guinea
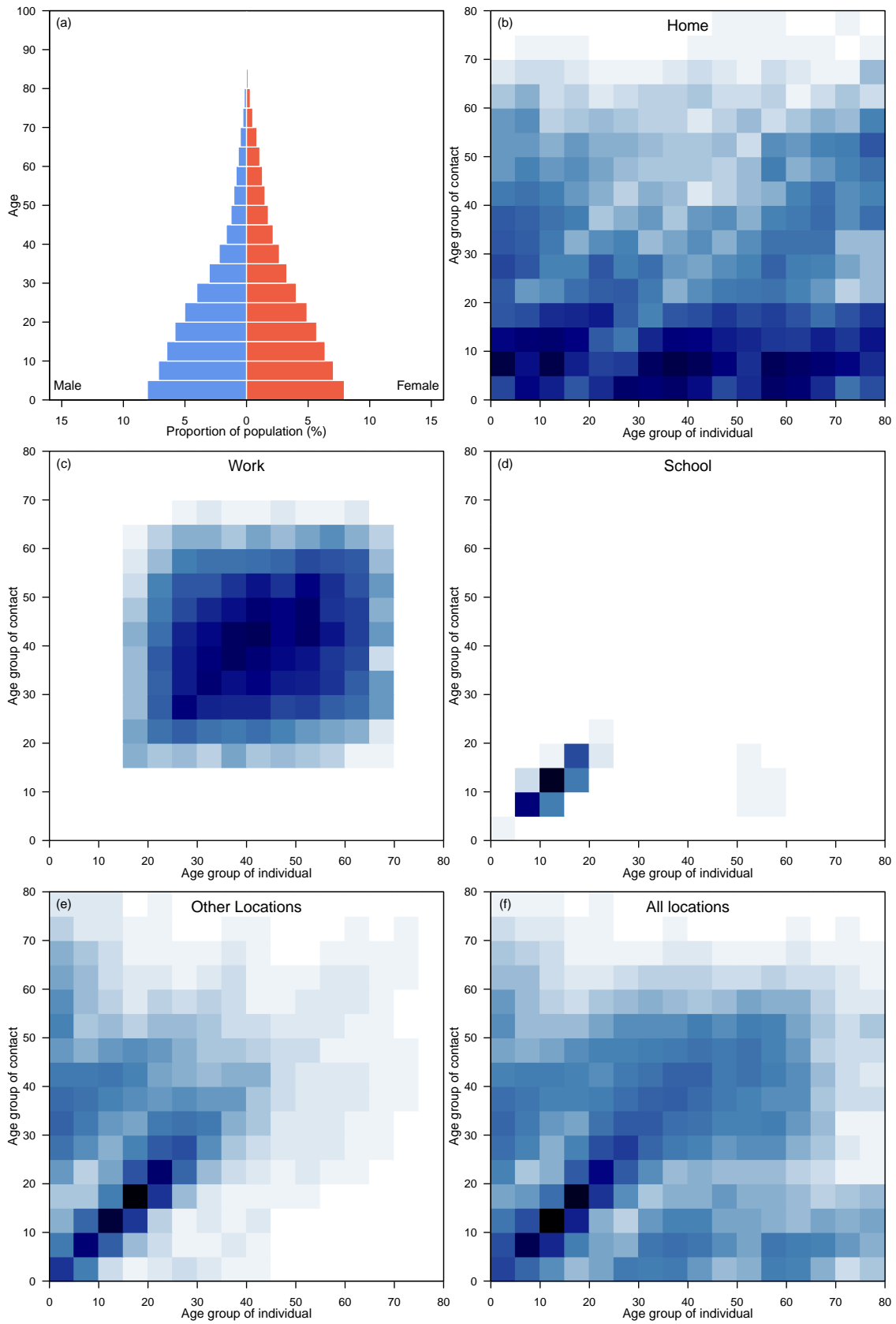
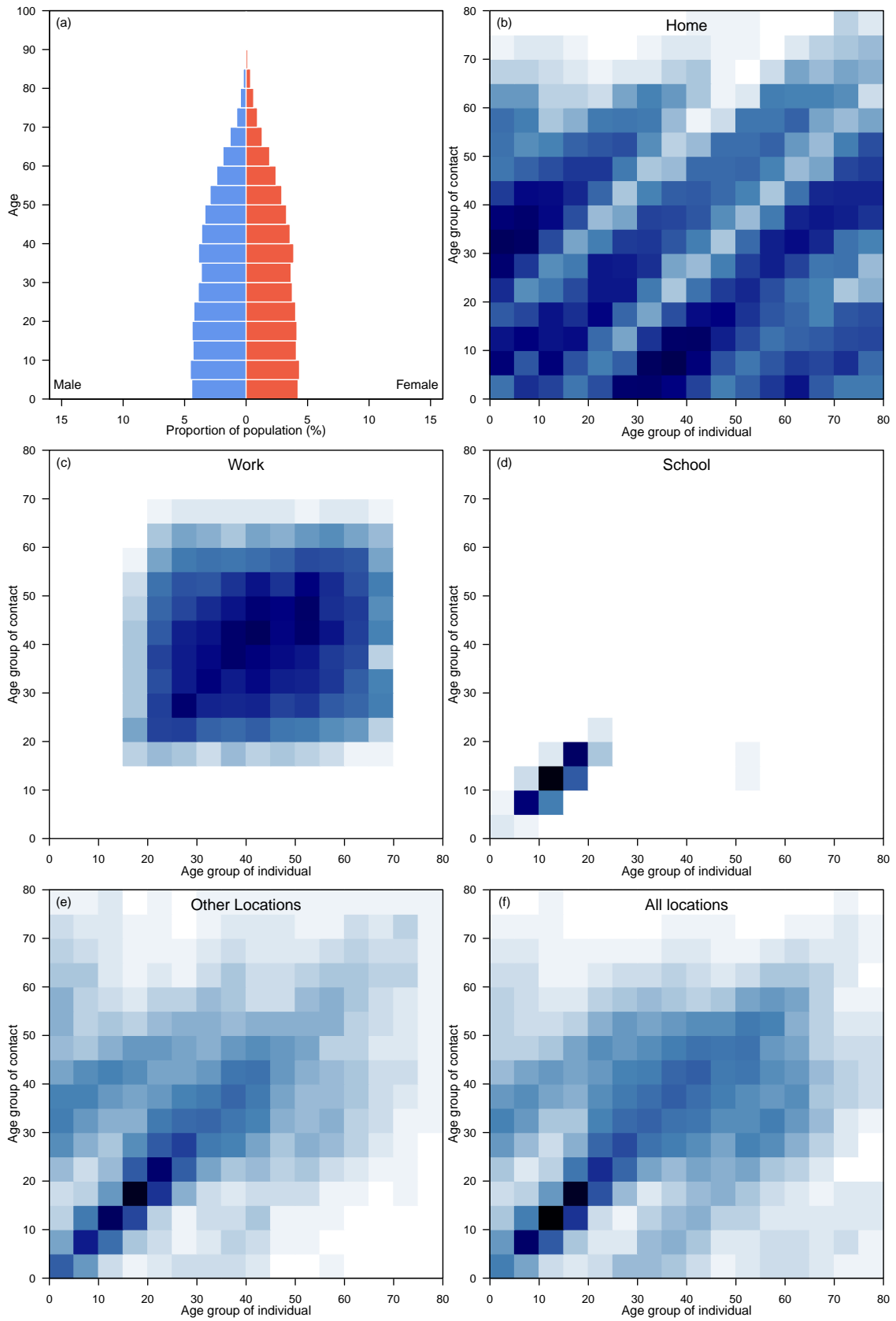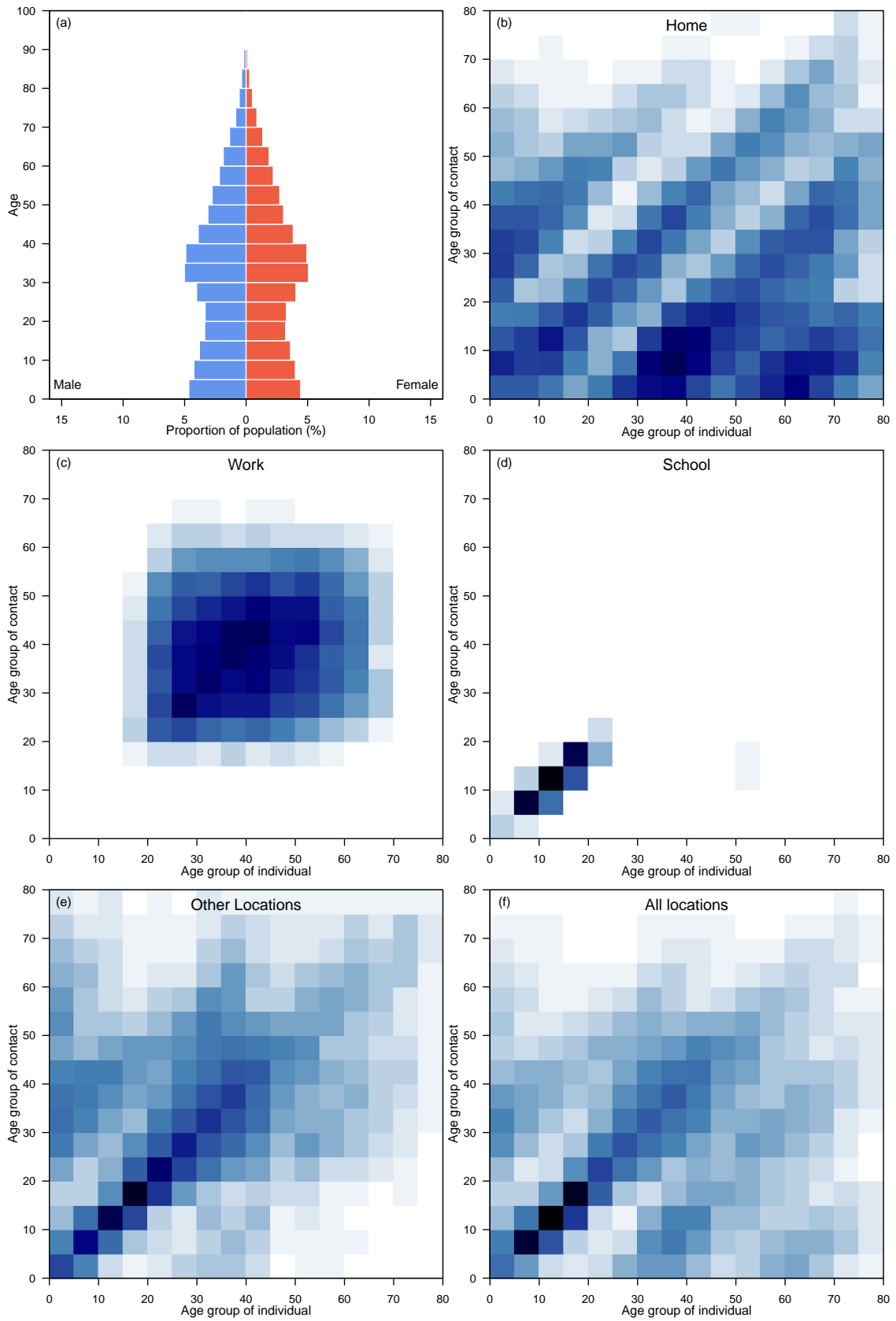# Guinea–Bissau
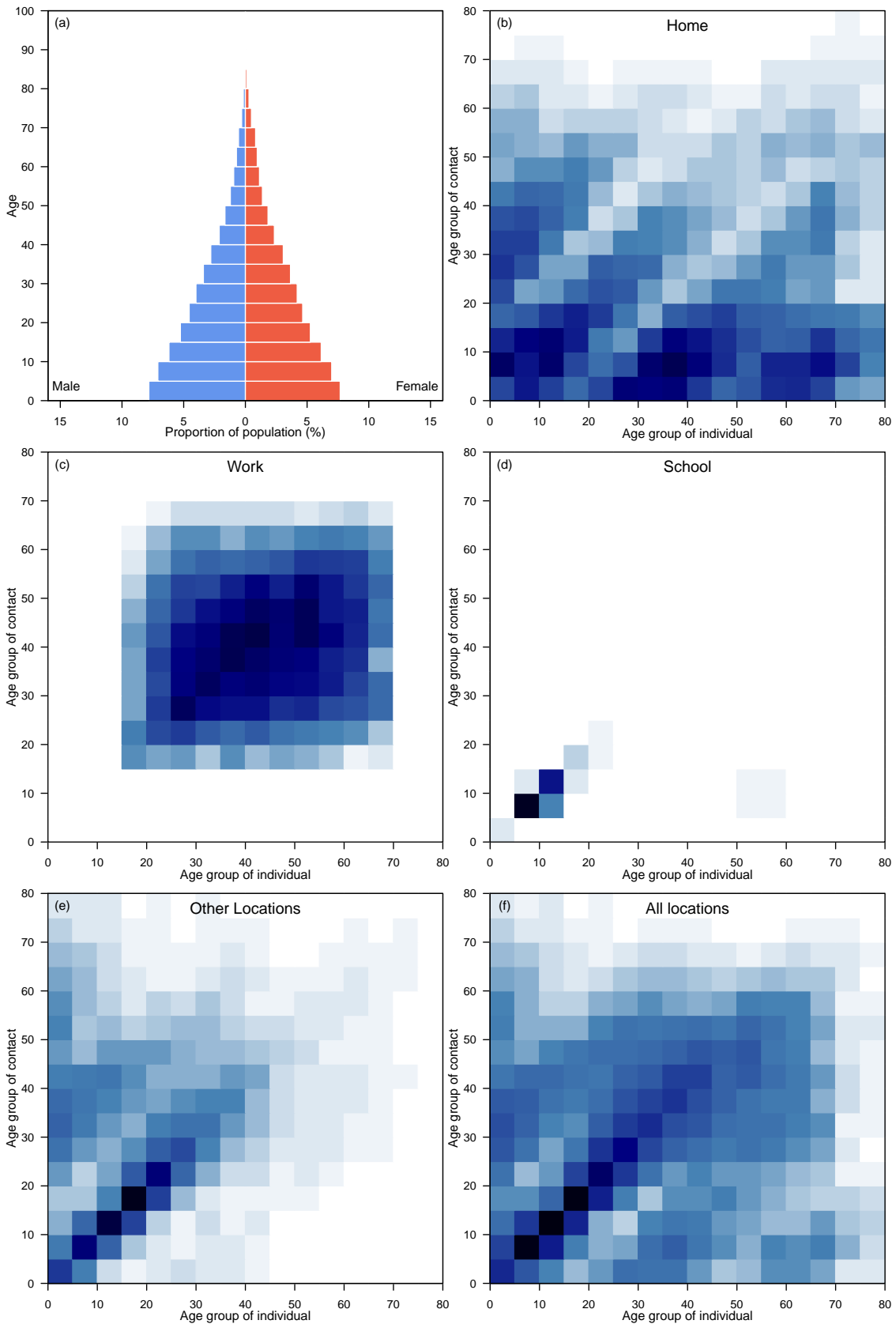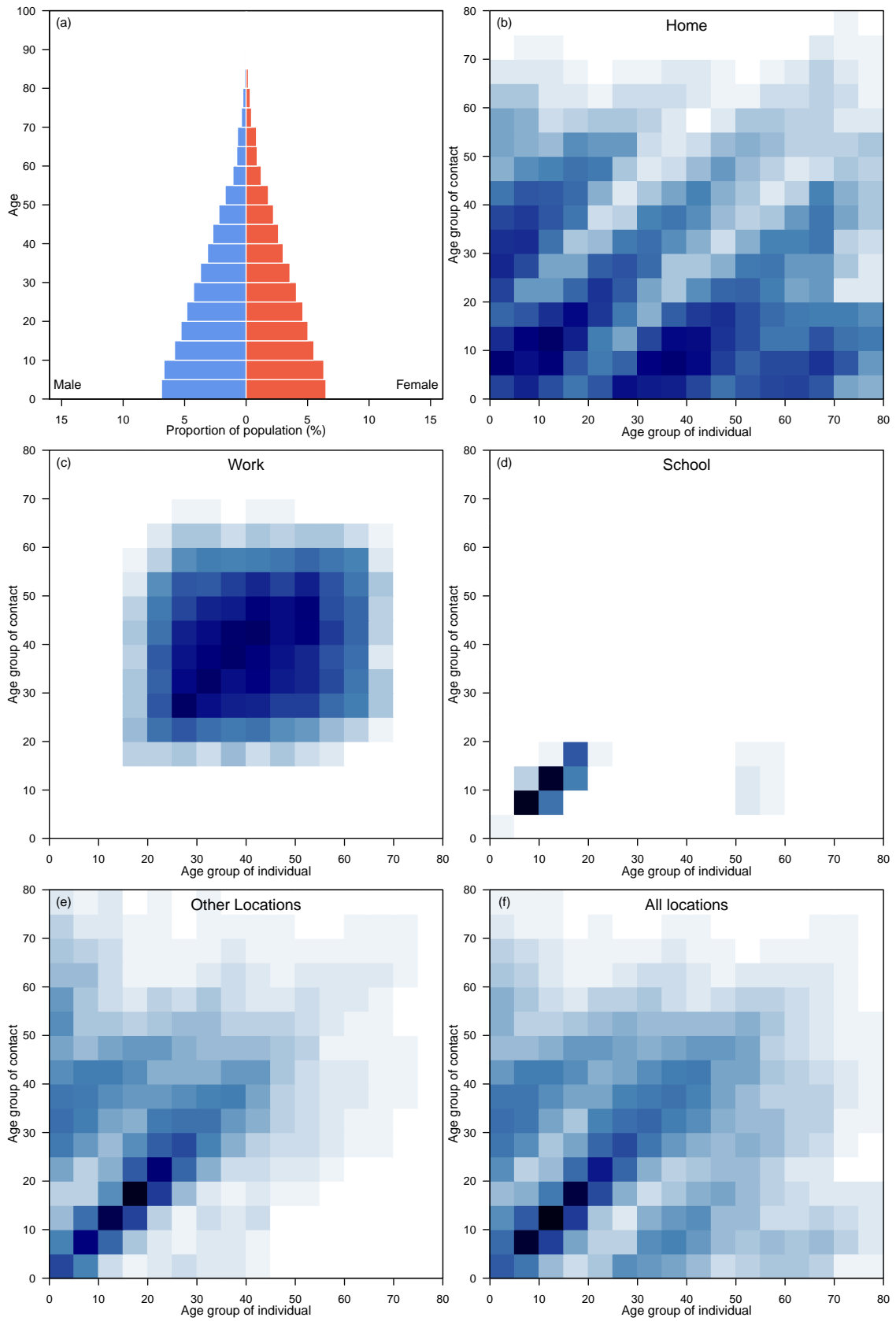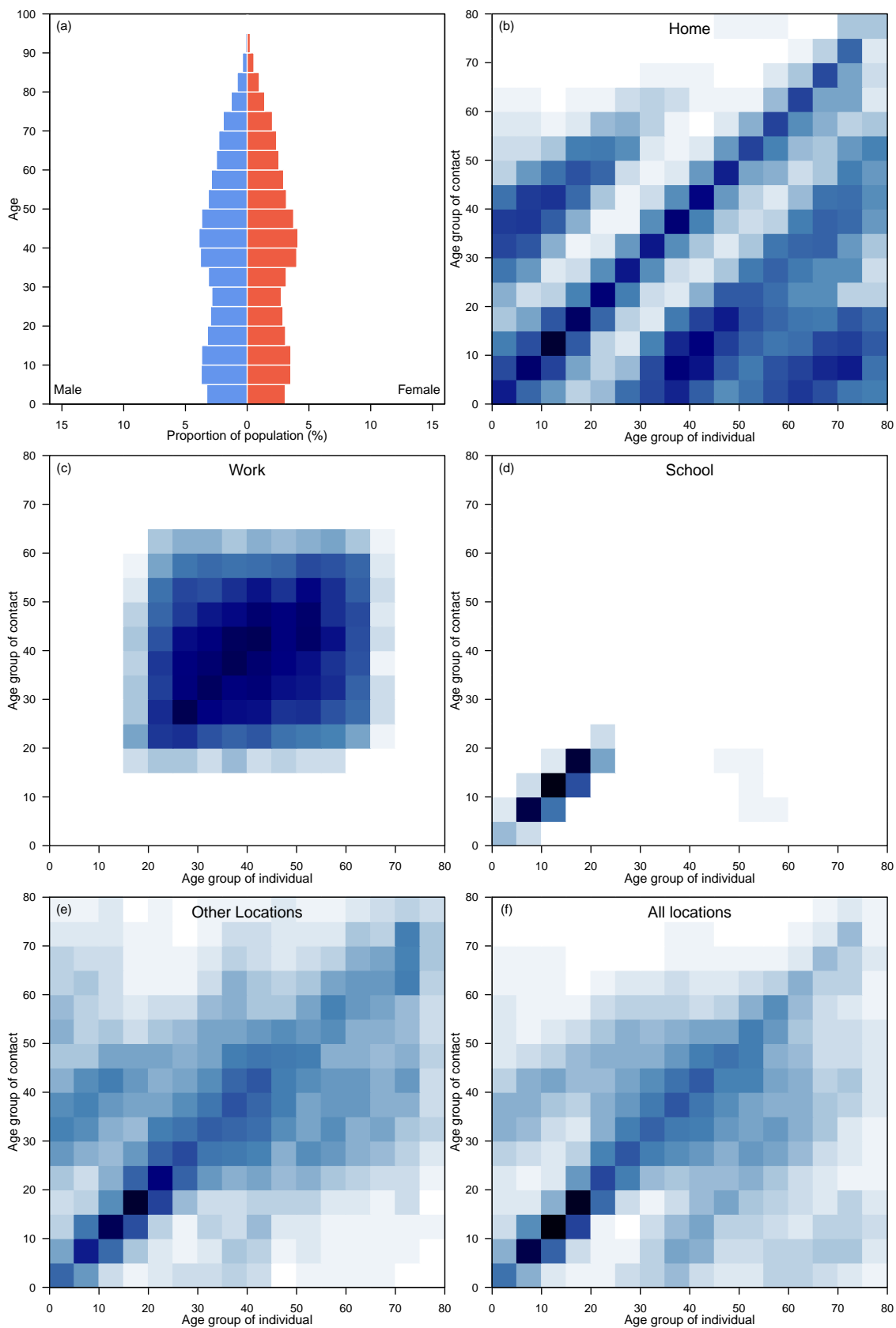
Guyana
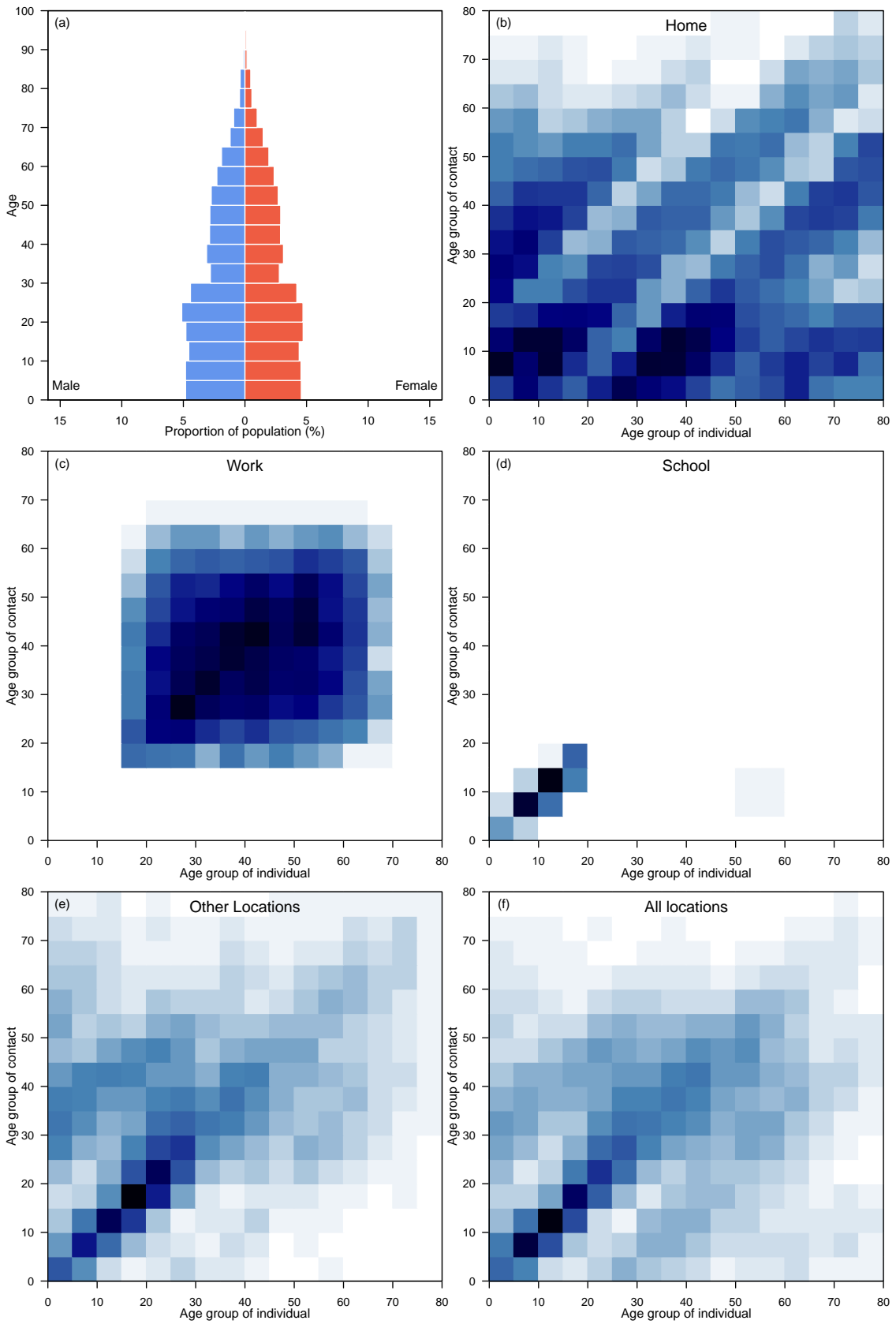
Honduras
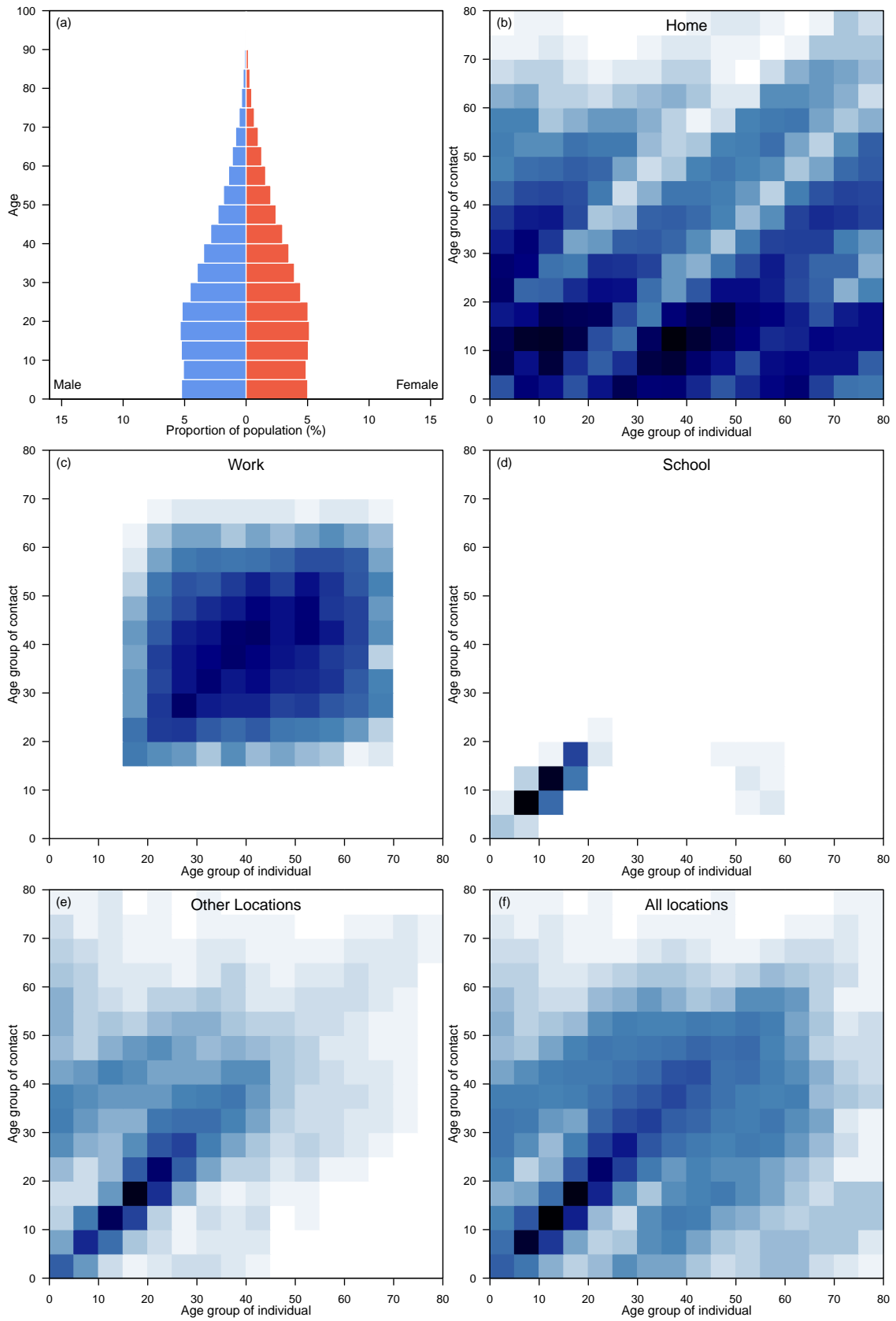
Hungary
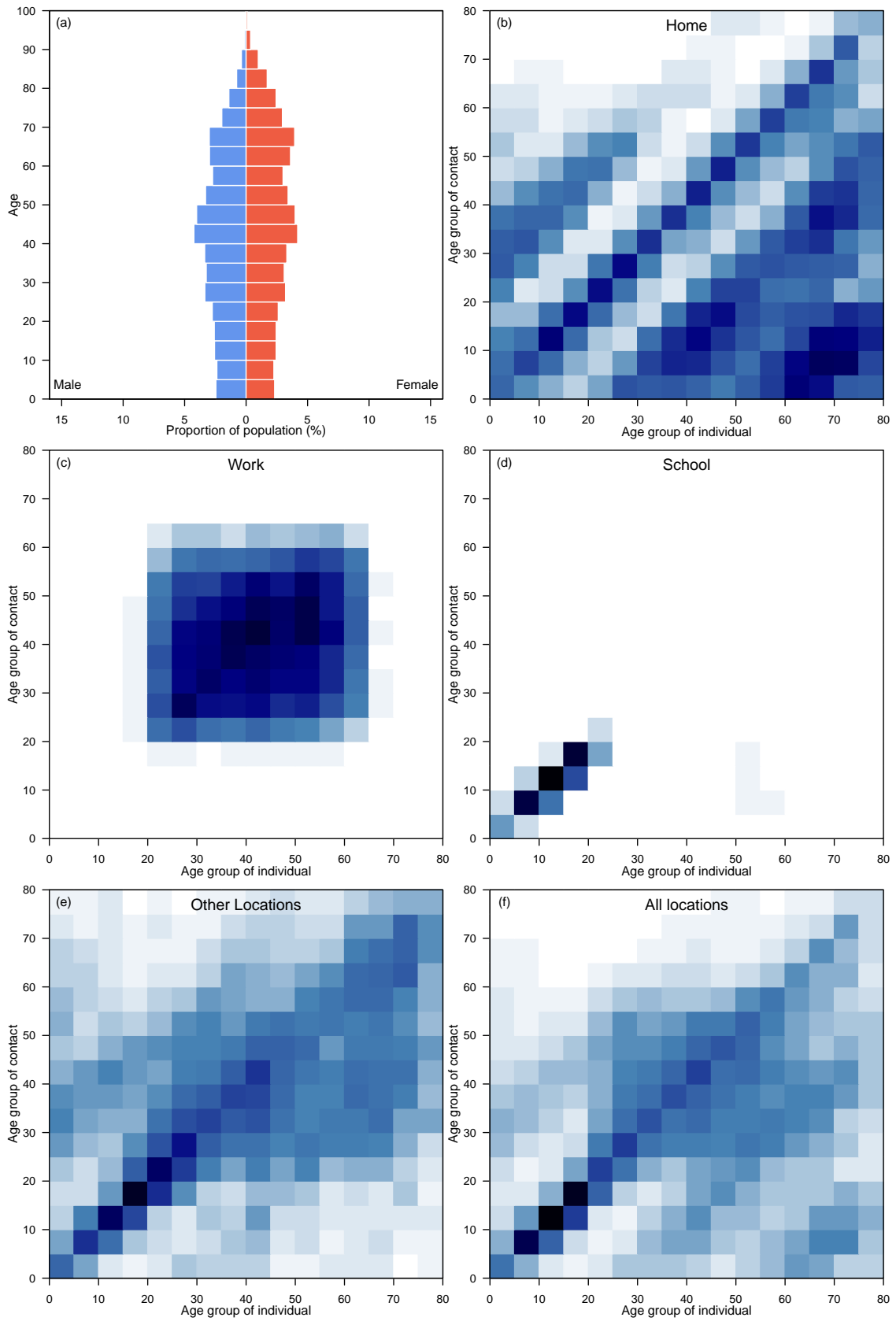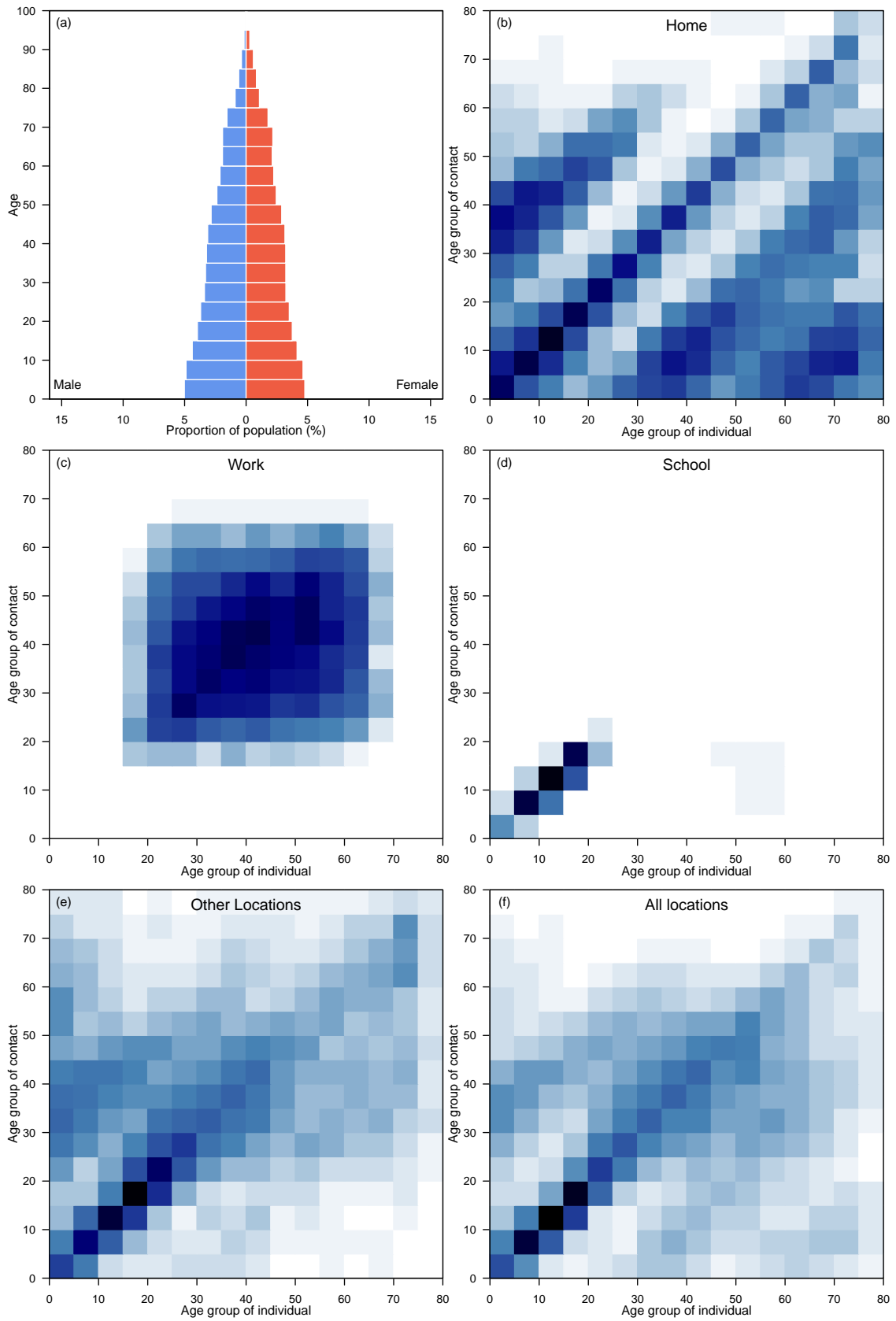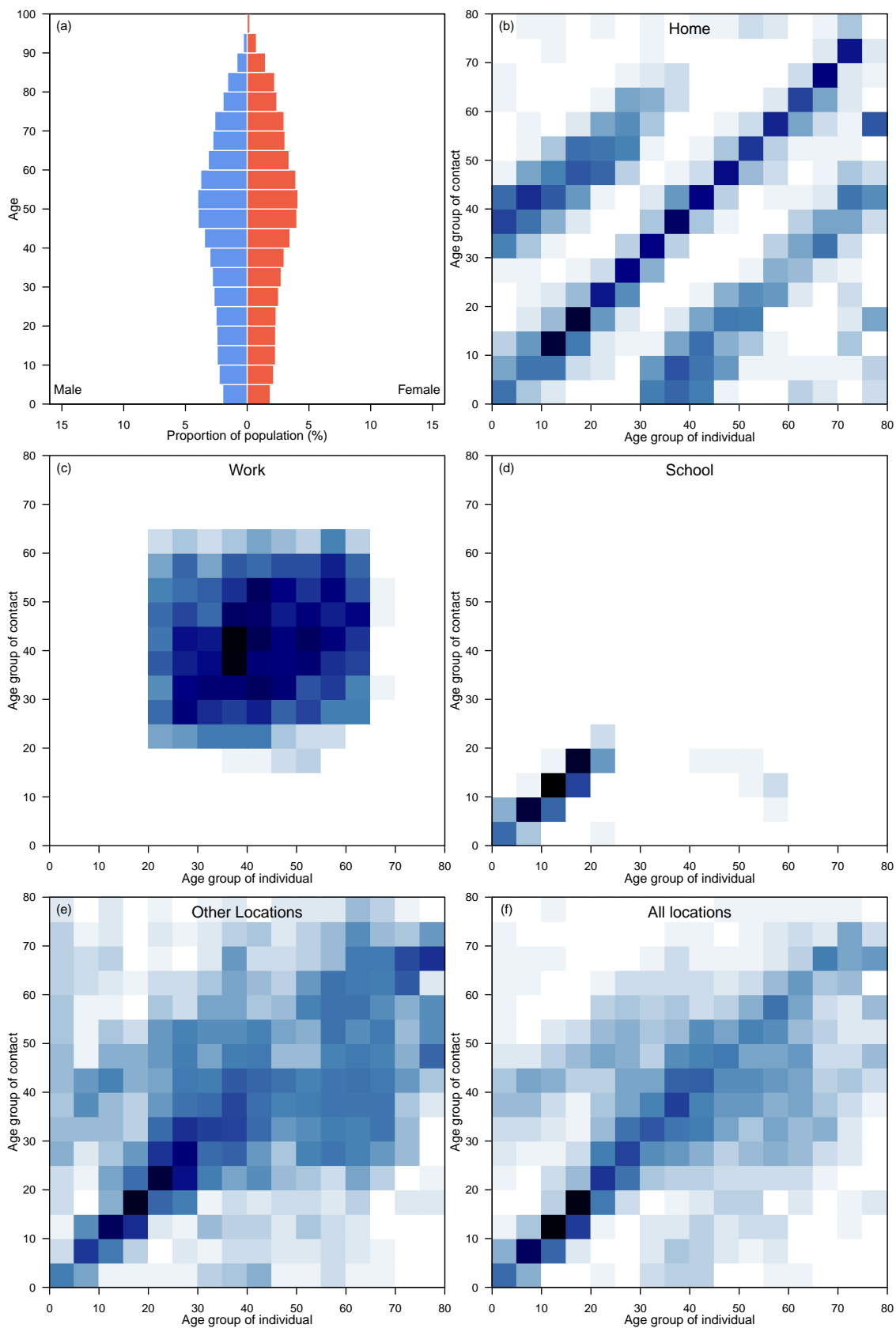
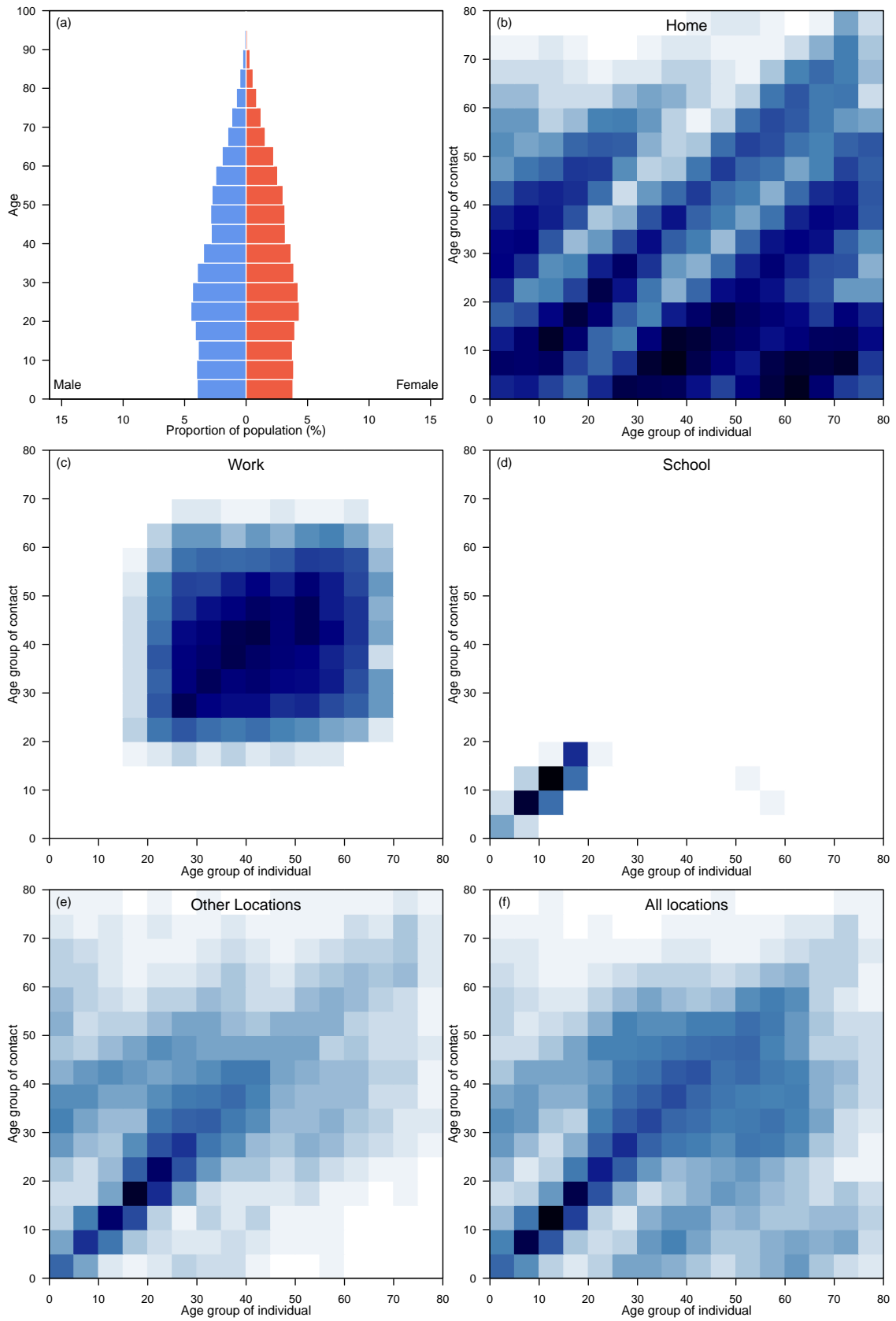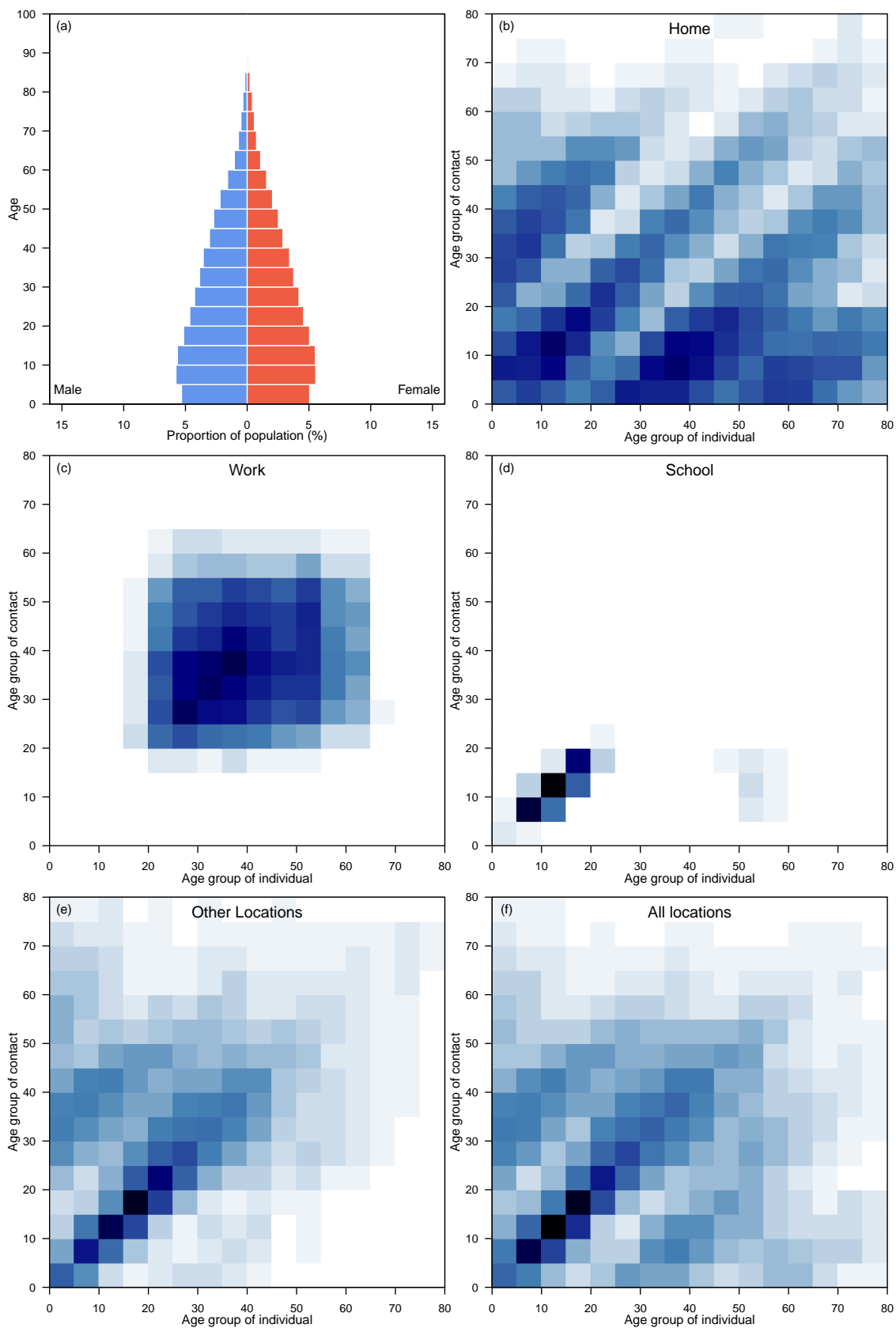# Iceland
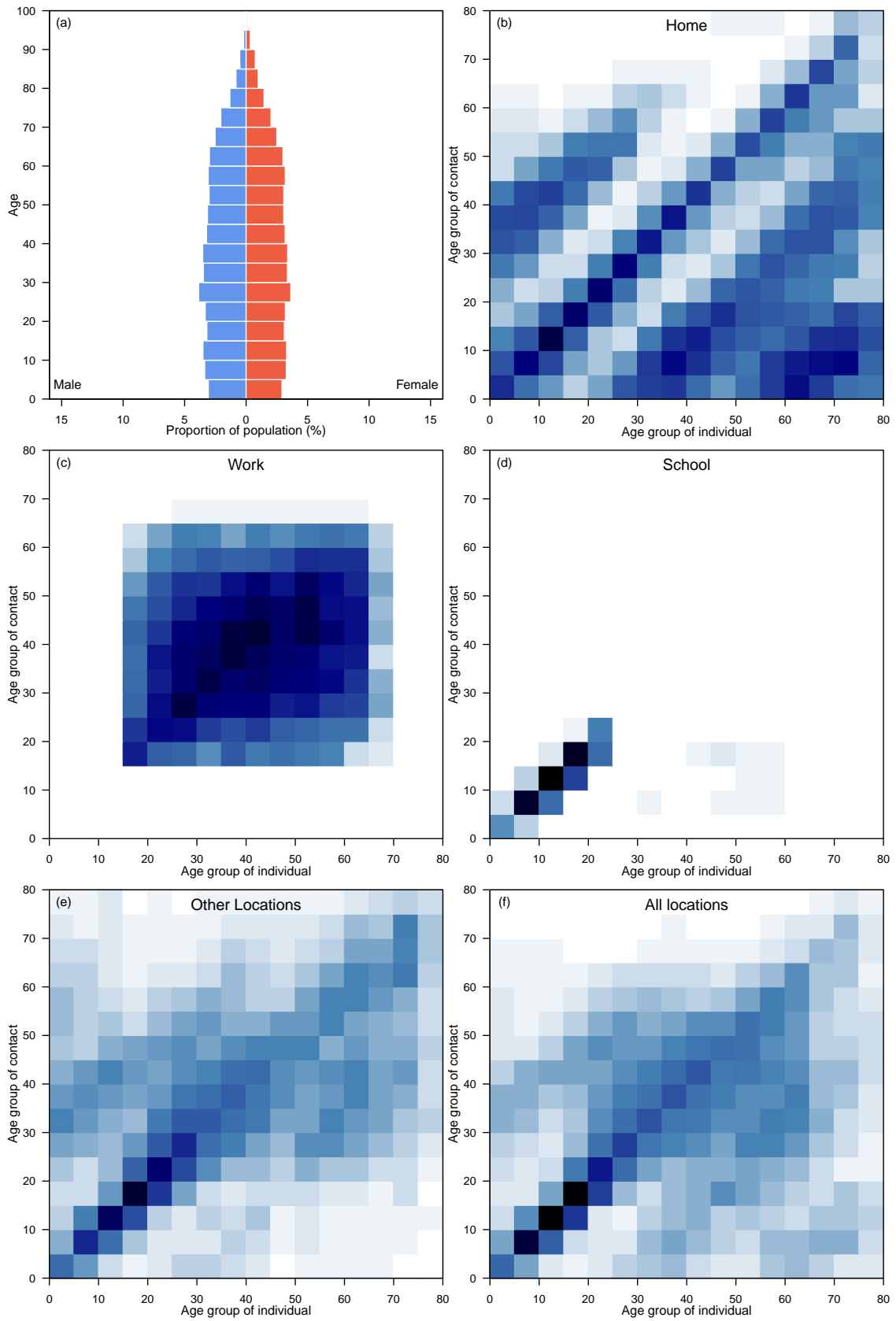
India
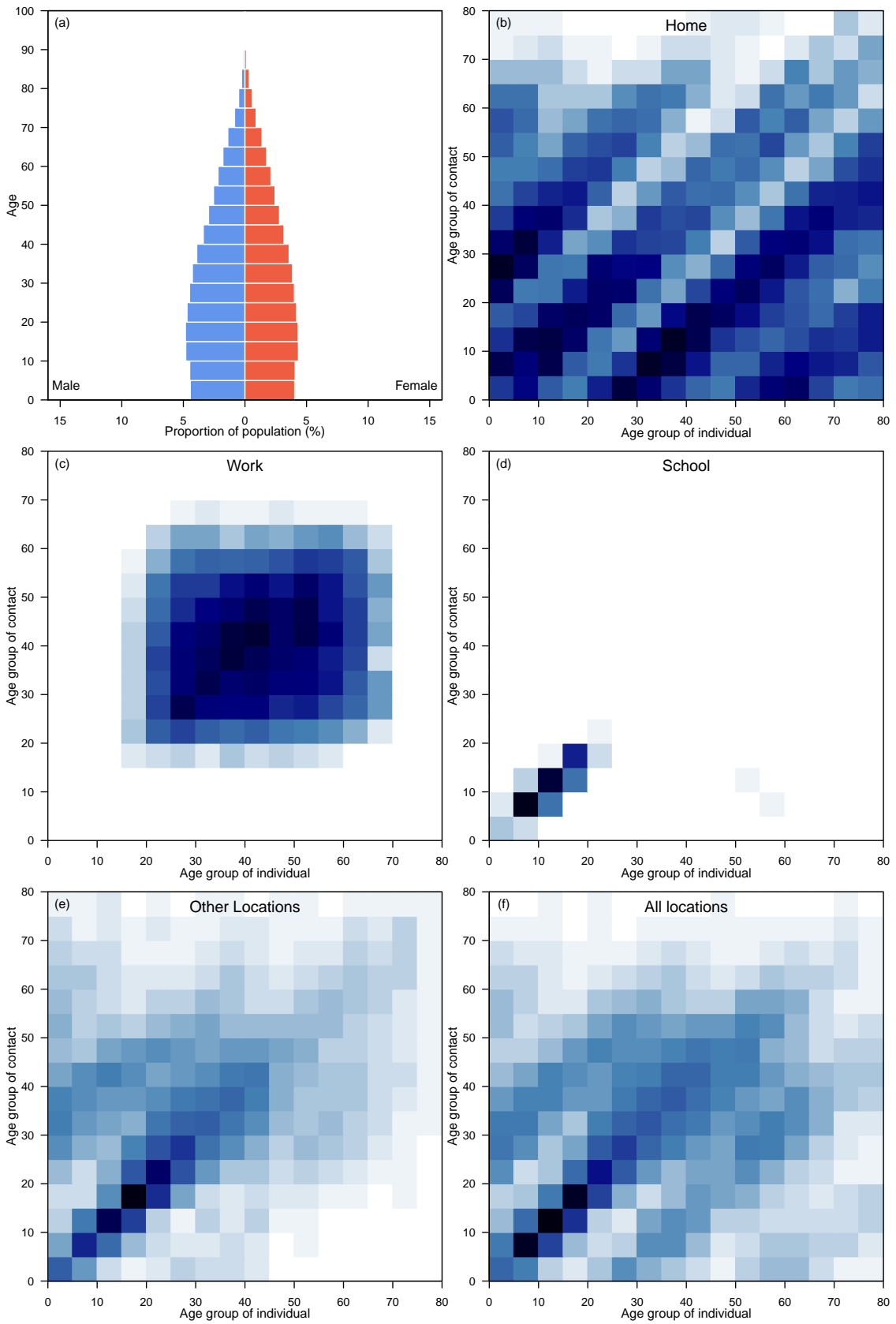
Indonesia
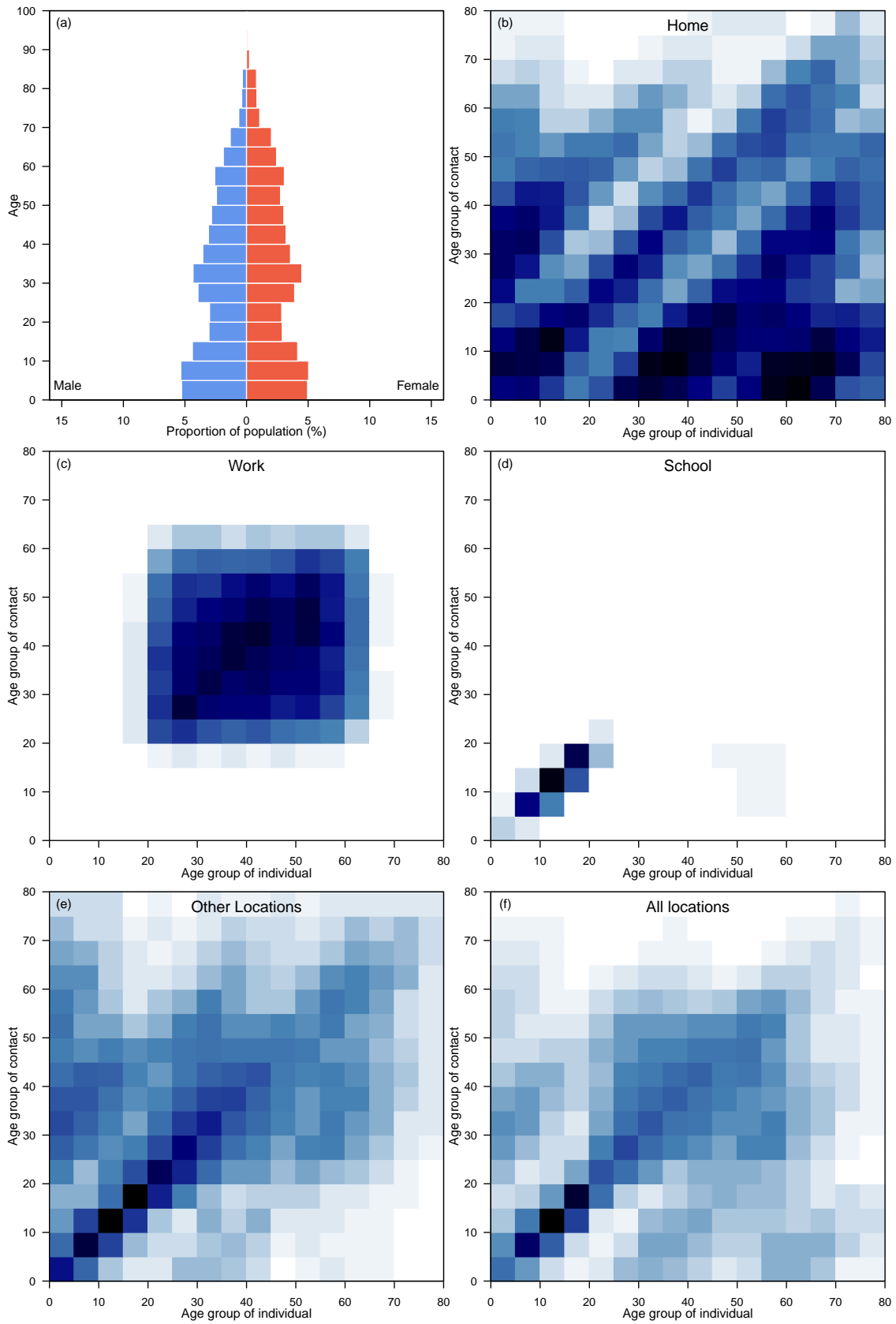
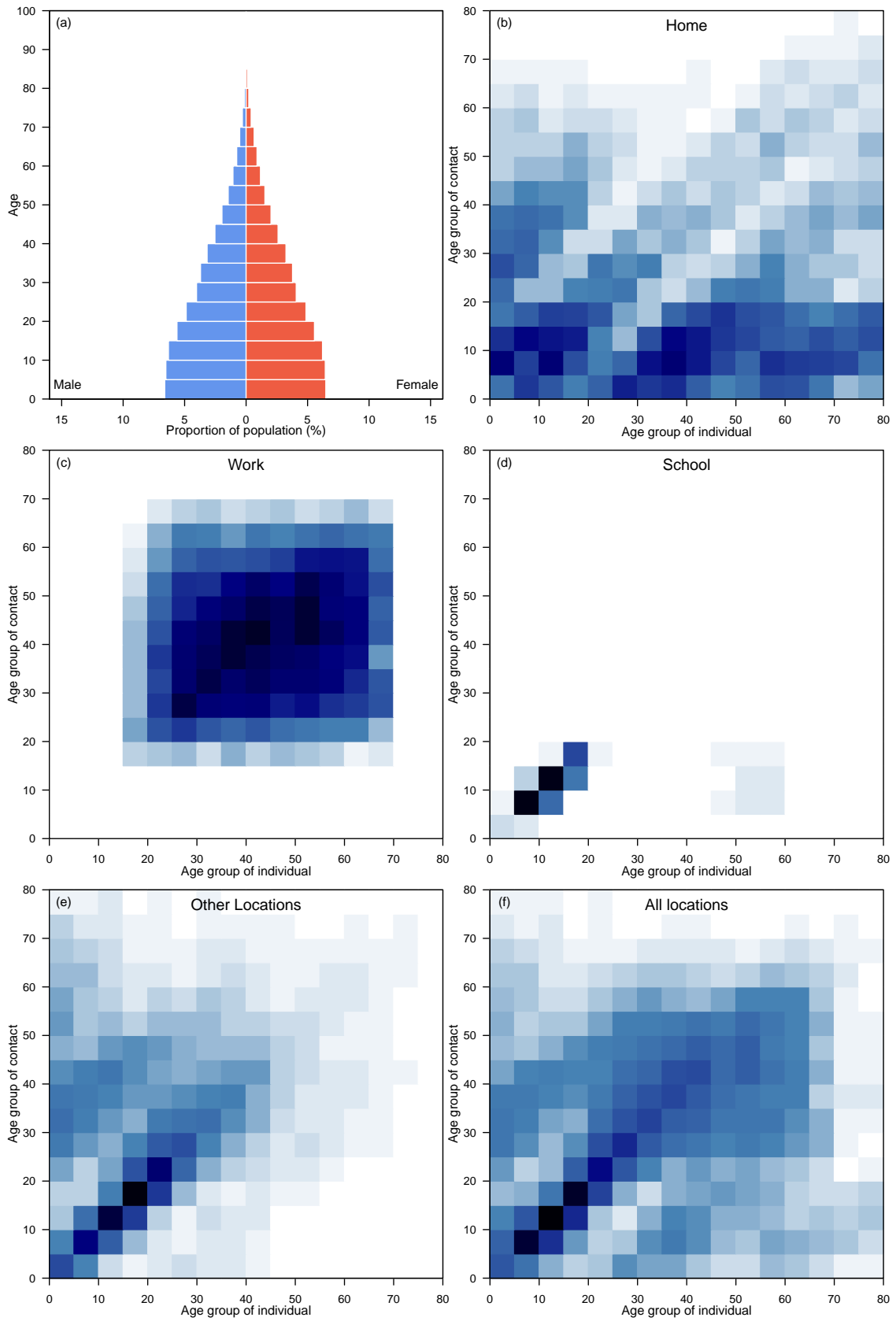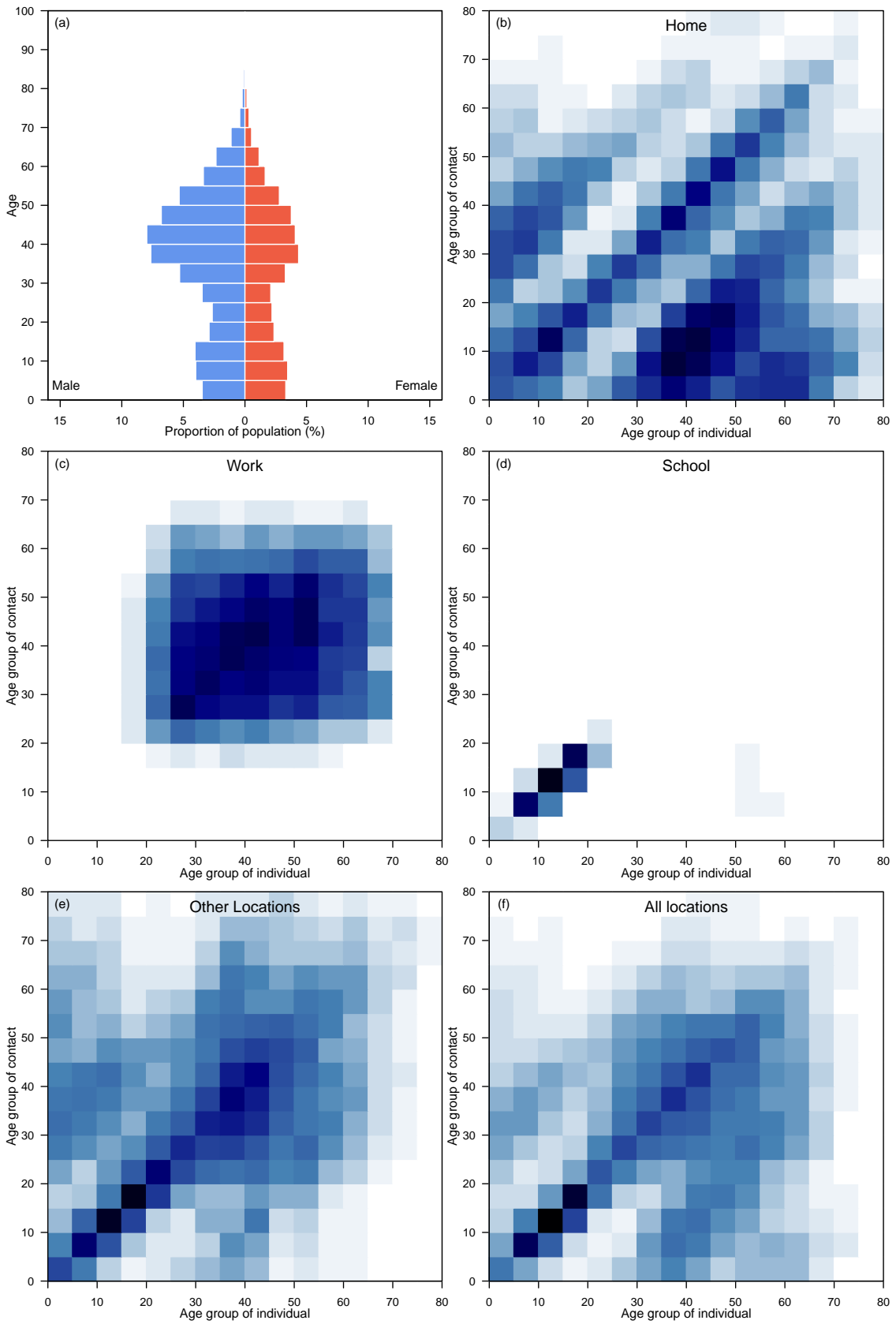# Iran (Islamic Republic of)

Iraq

Ireland

Israel
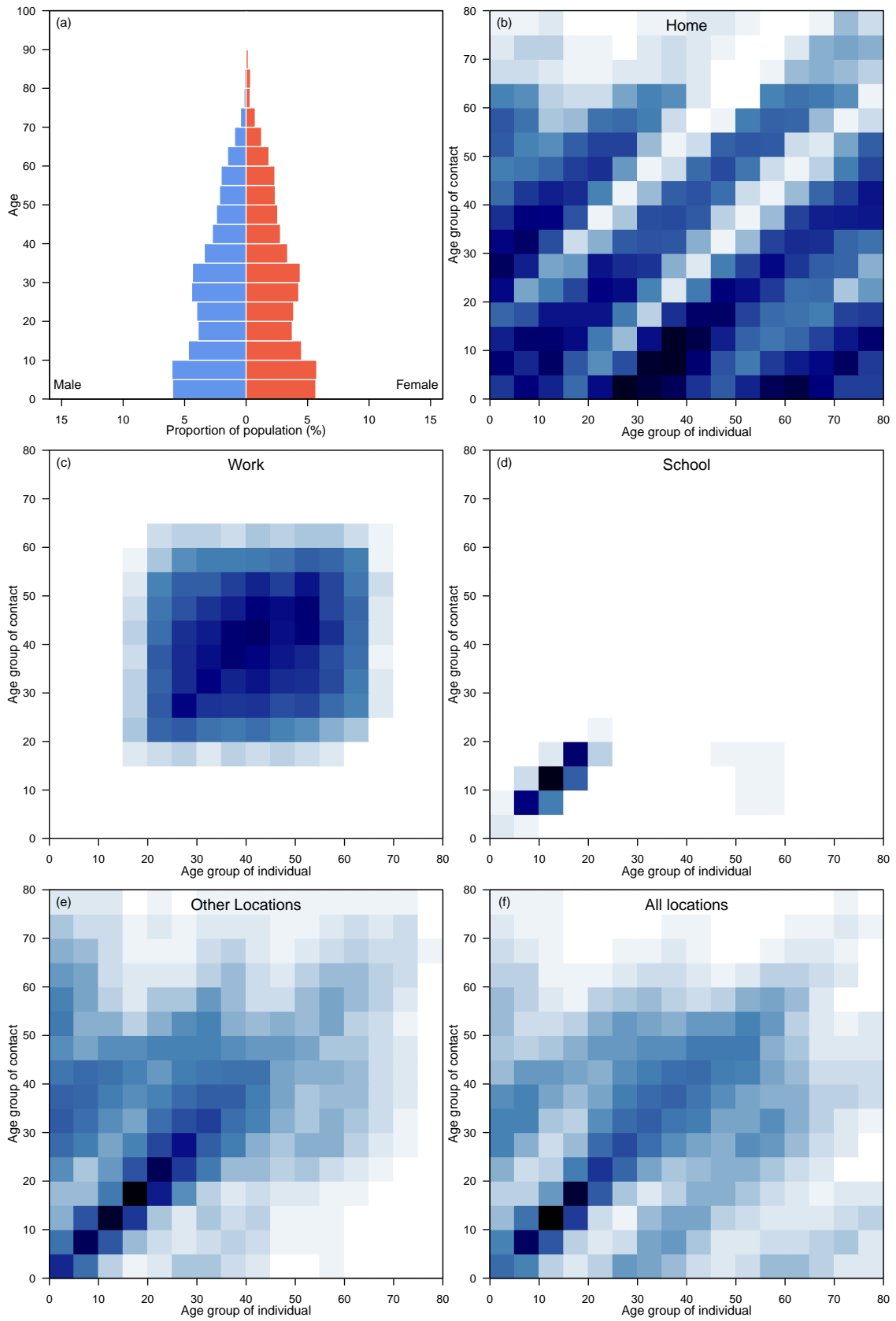
Italy

Jamaica

Jordan

Kazakhstan

Kenya
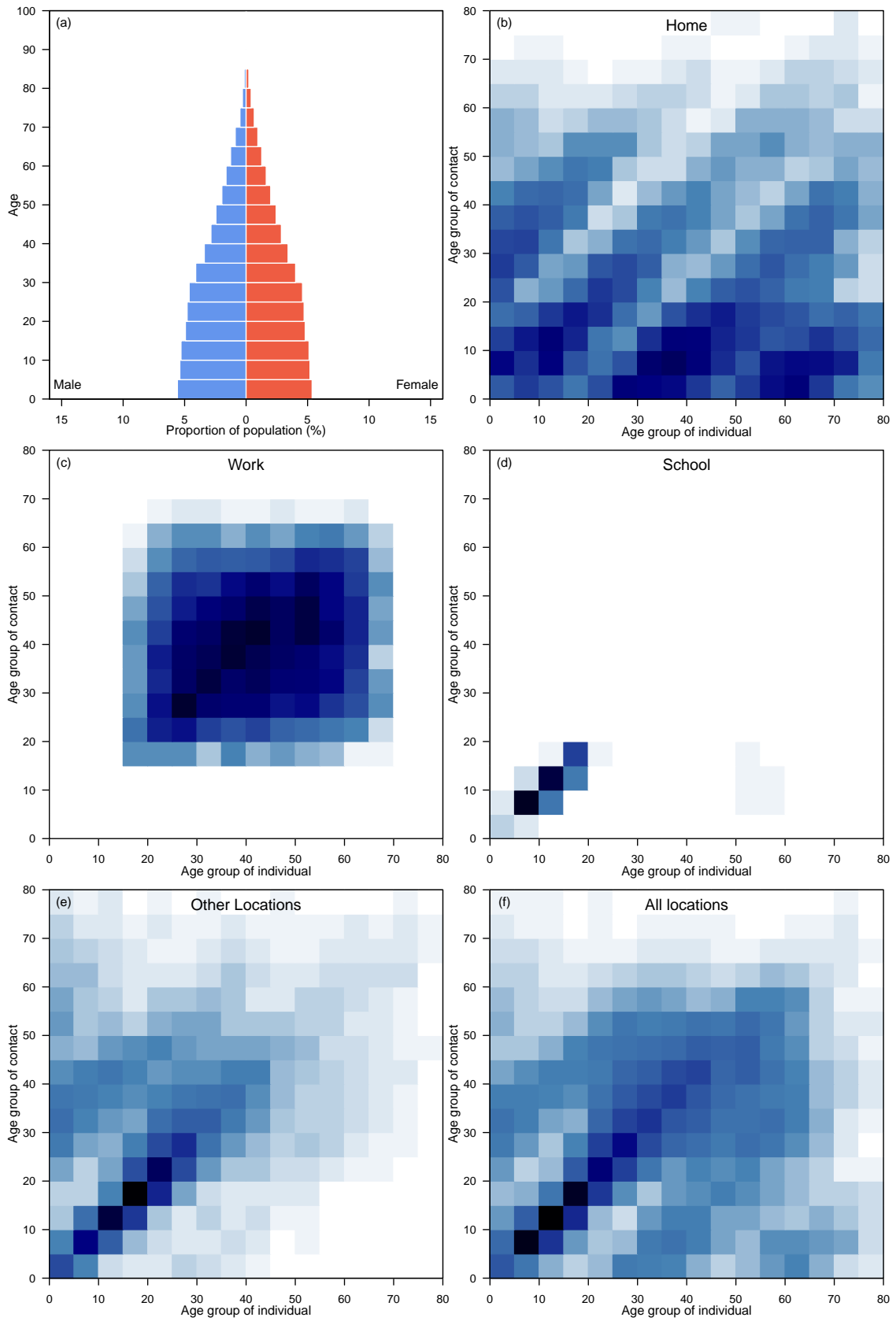
Kuwait

Kyrgyzstan

Lao People's Democratic Republic

Latvia
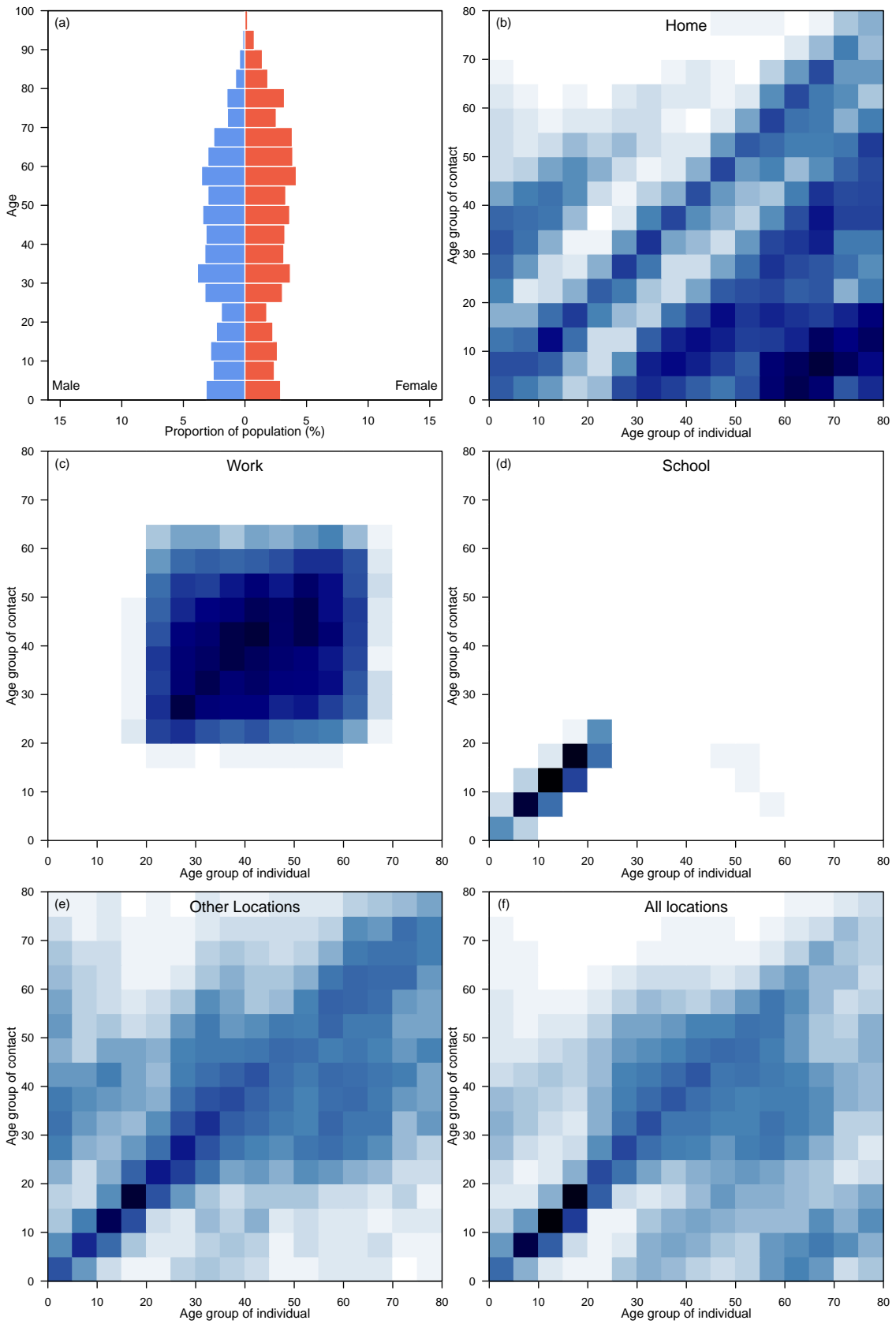
Lesotho
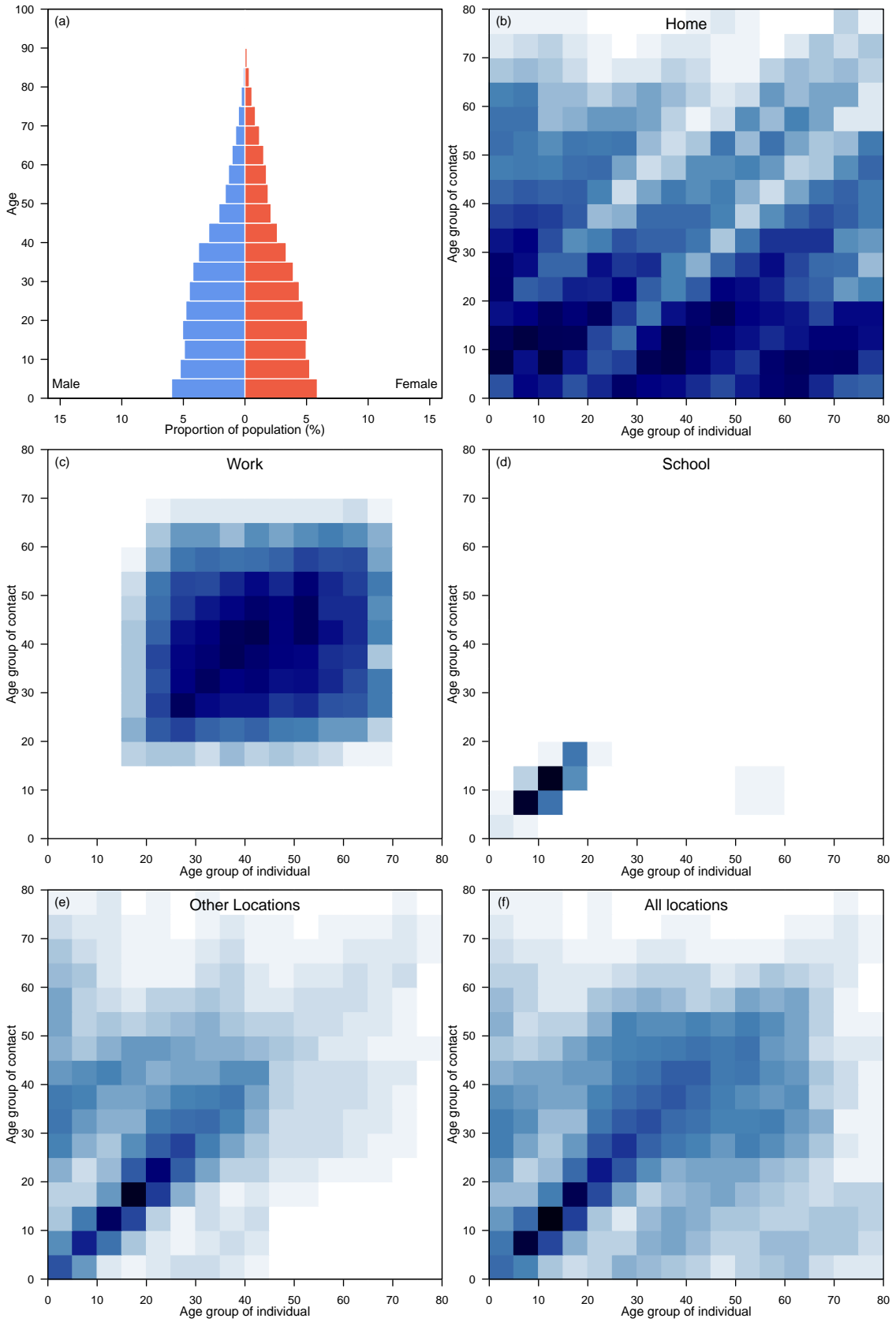
Liberia
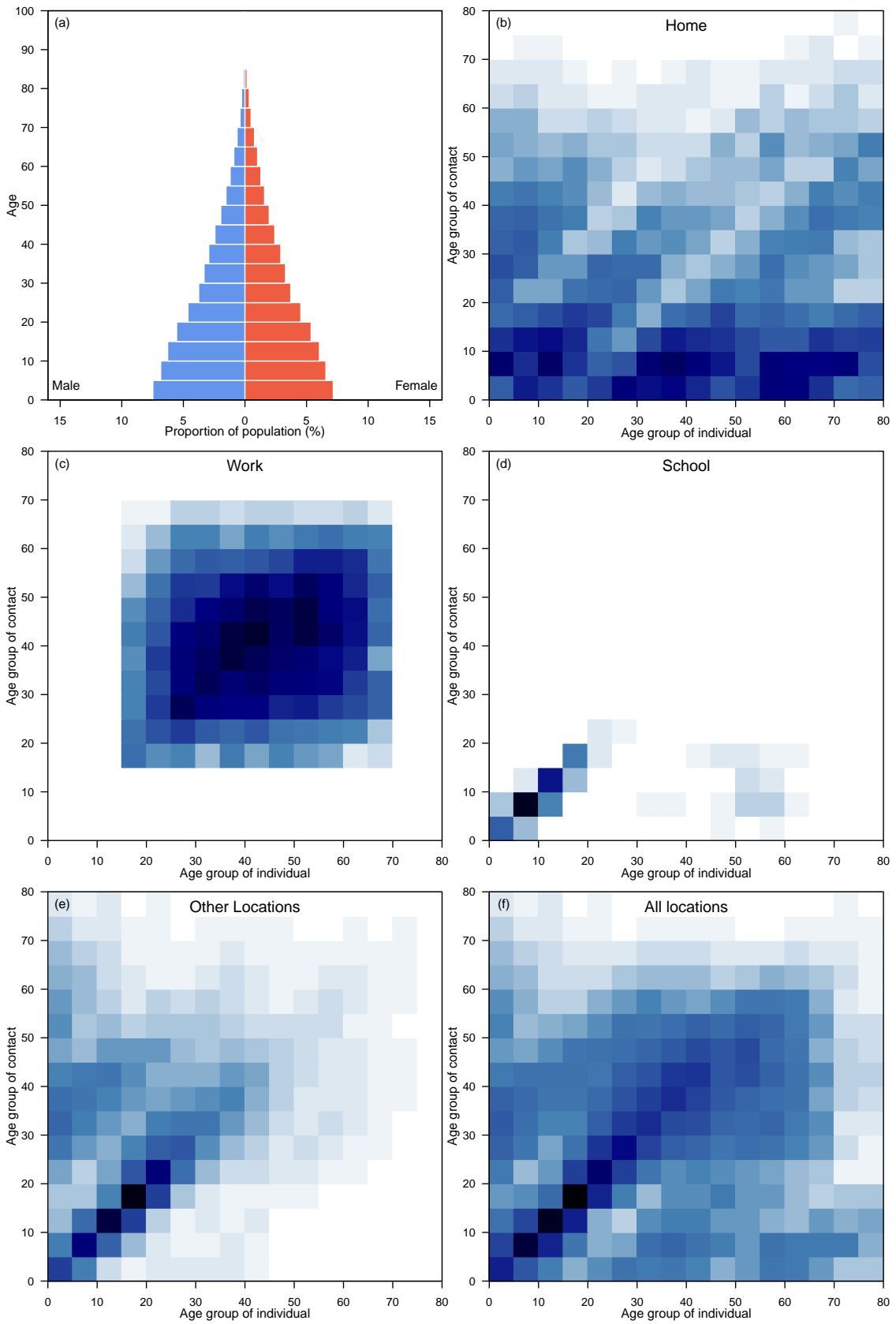
Libya
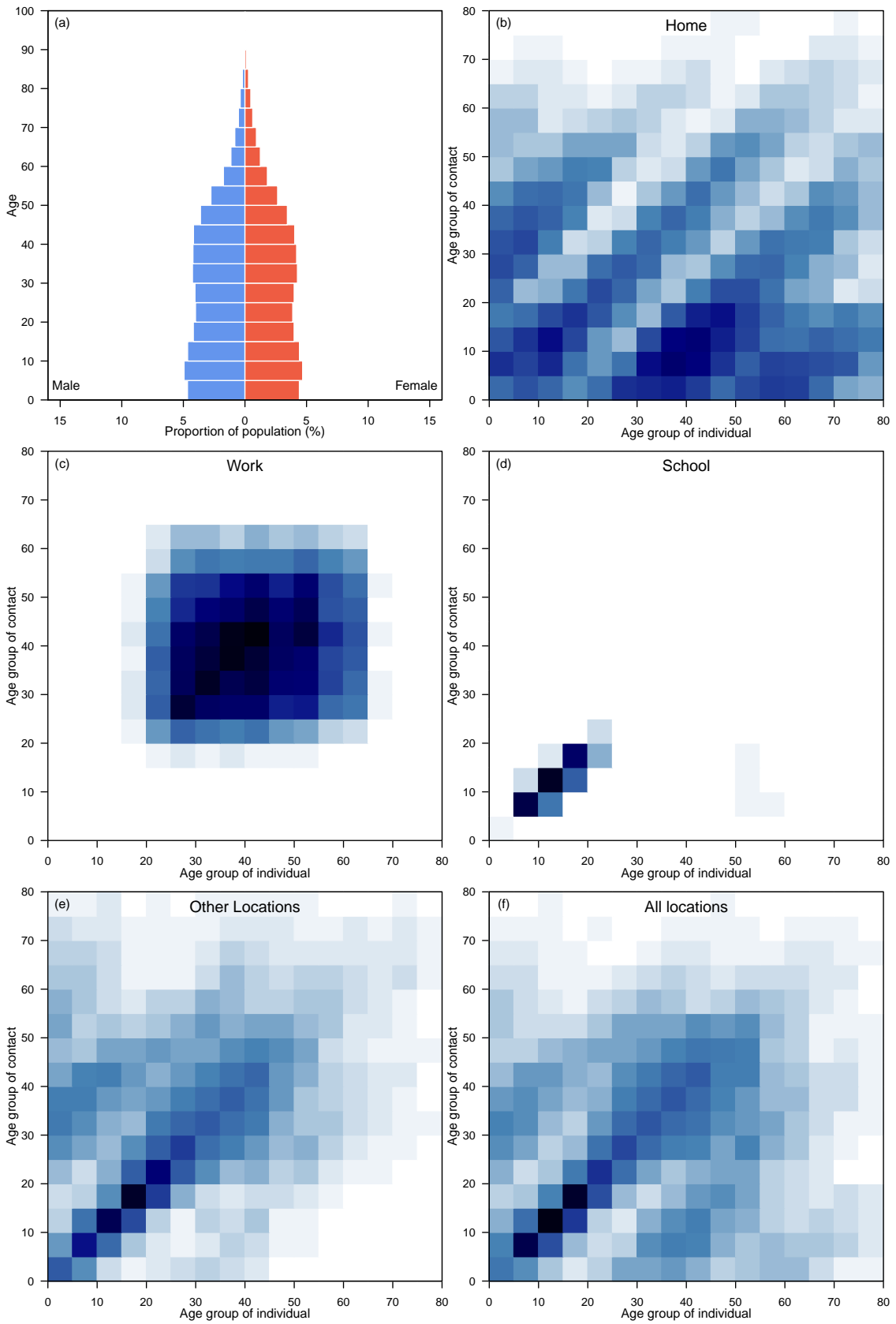
Lithuania
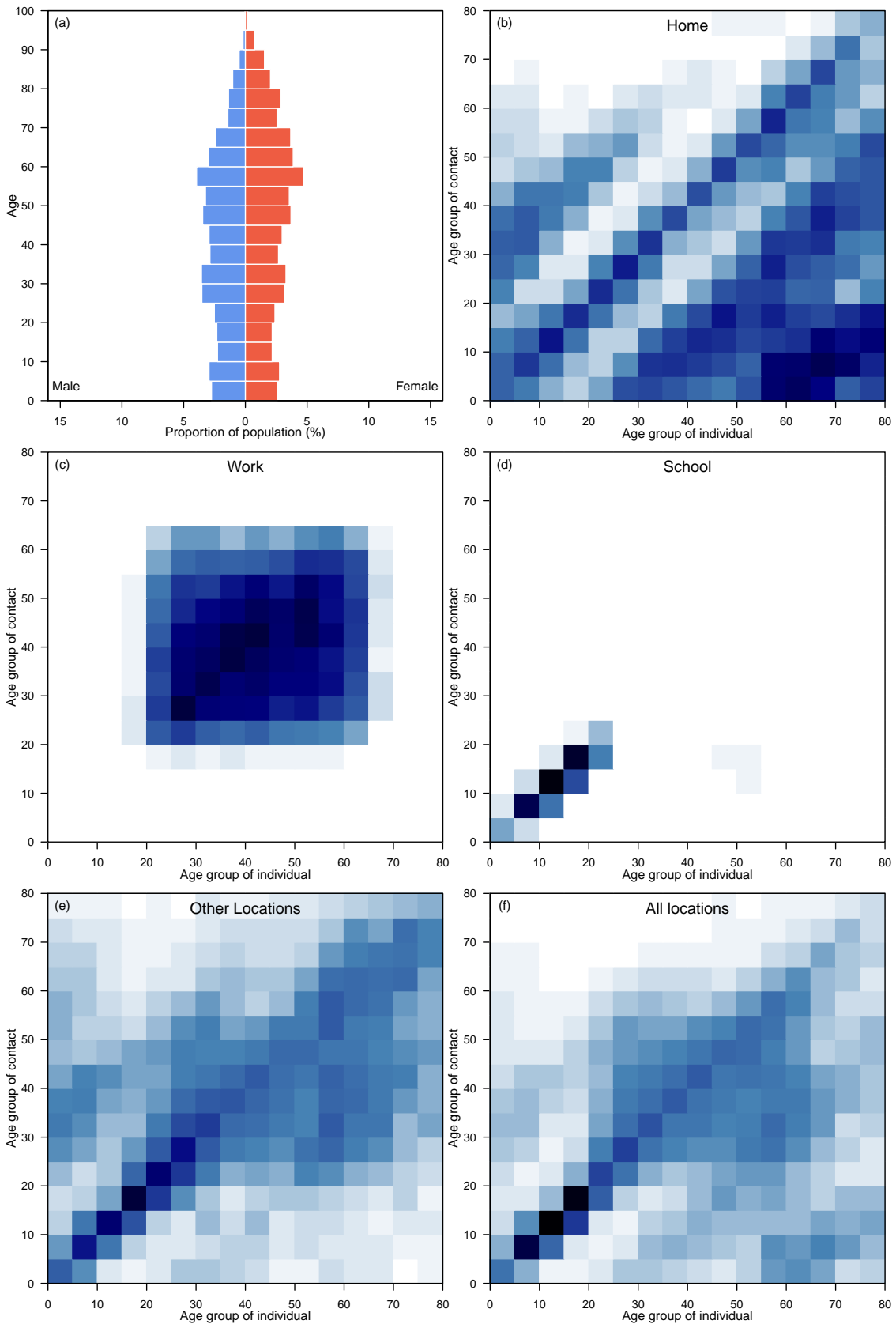
Luxembourg
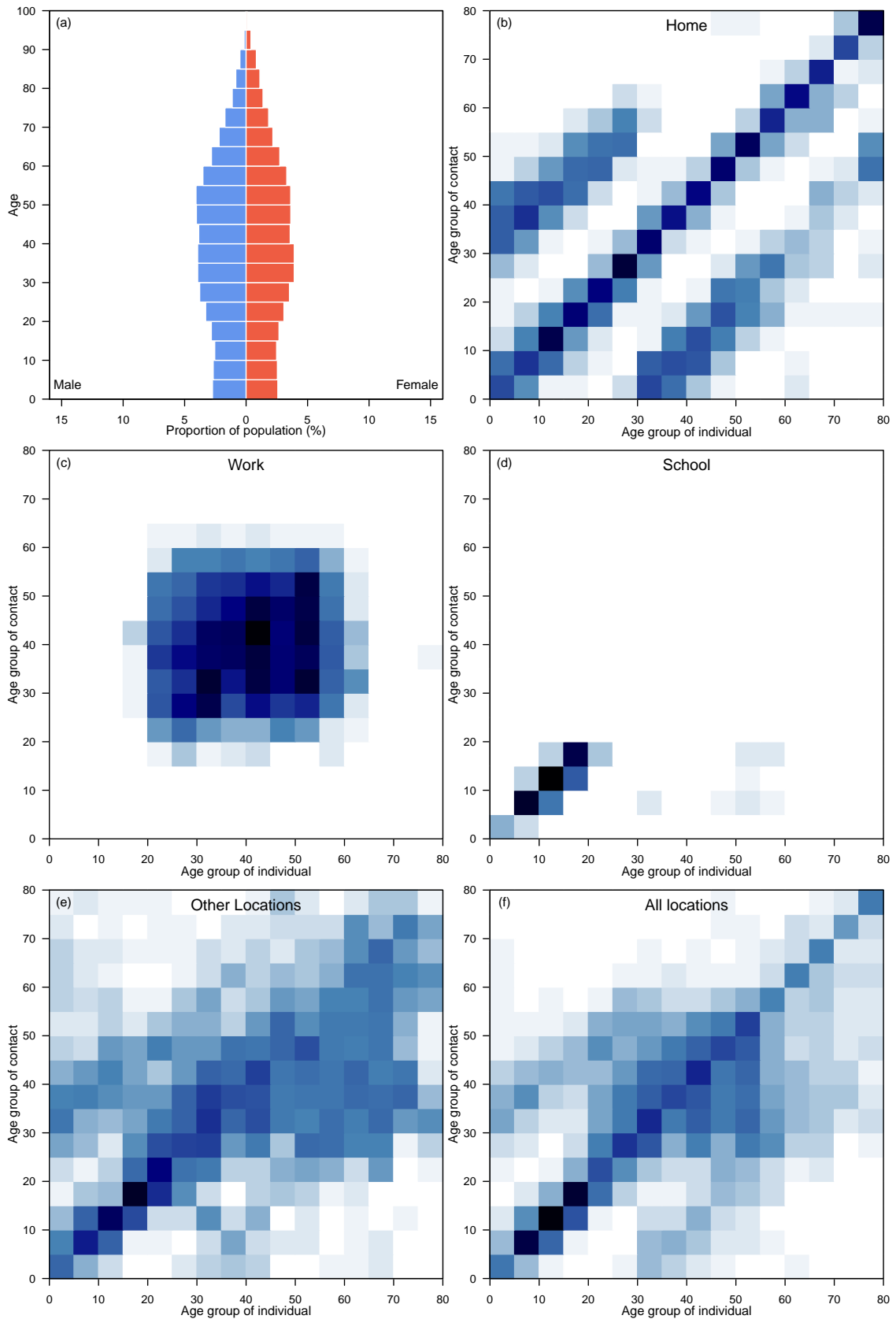
Madagascar
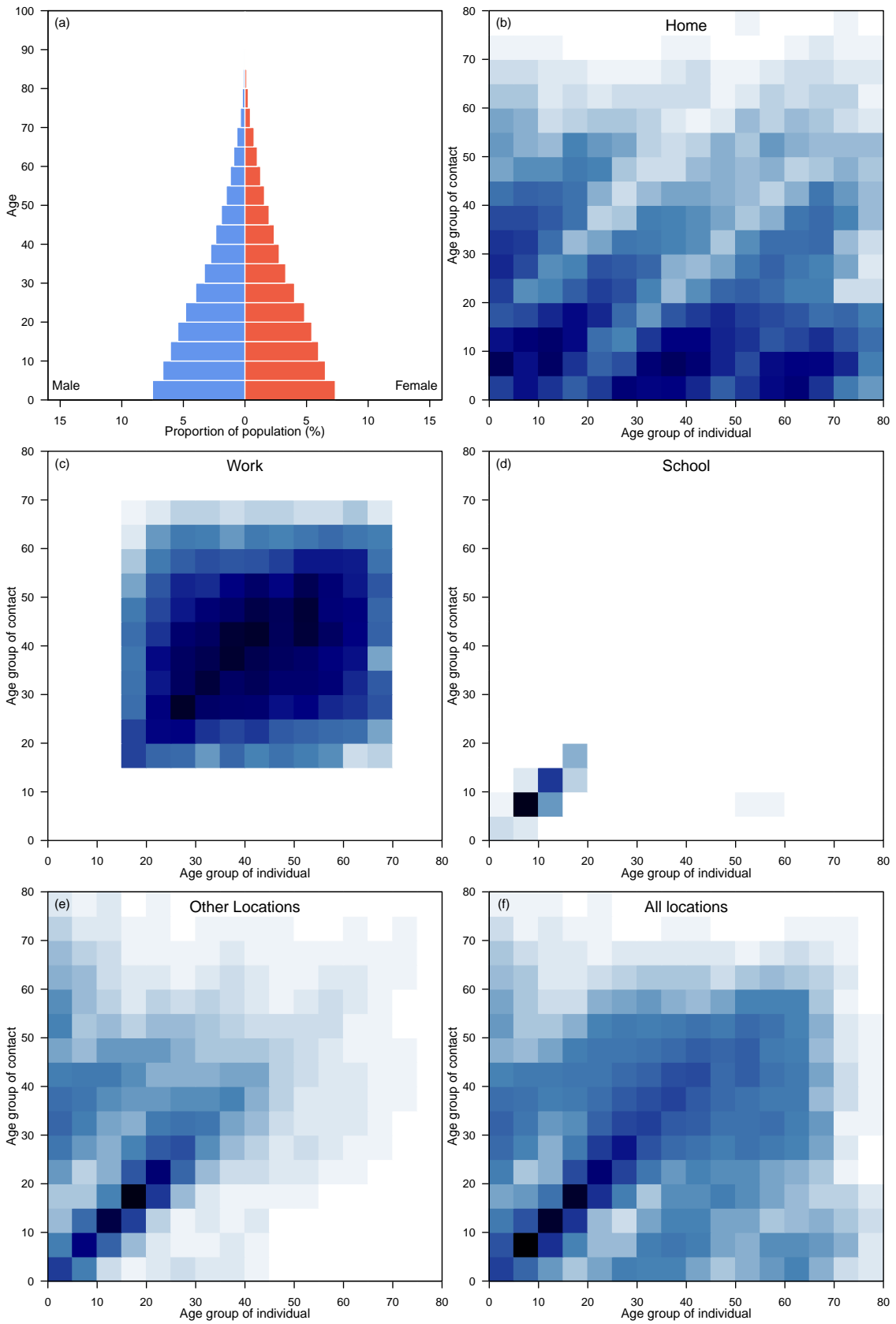
Malawi
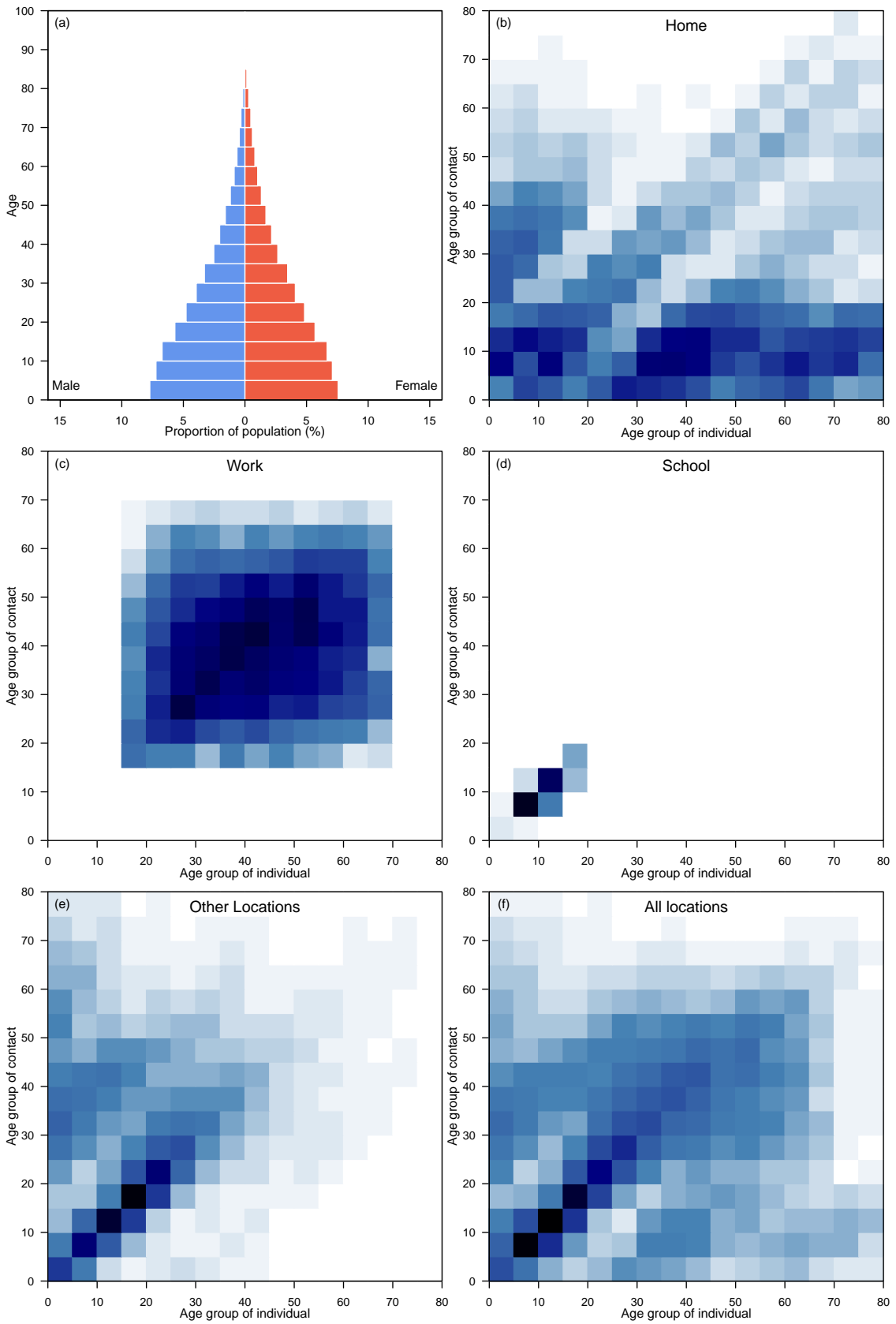
Malaysia
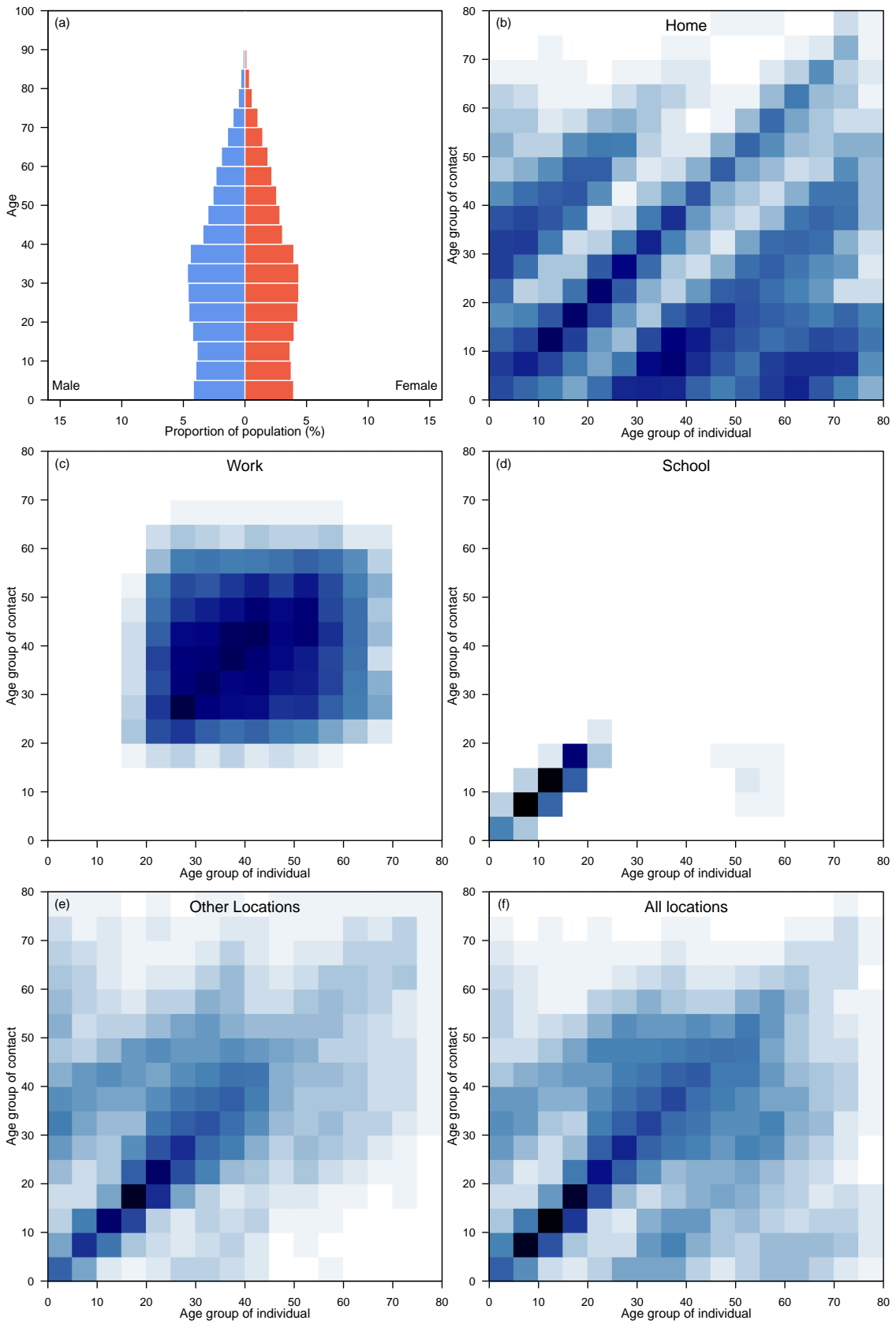
Maldives
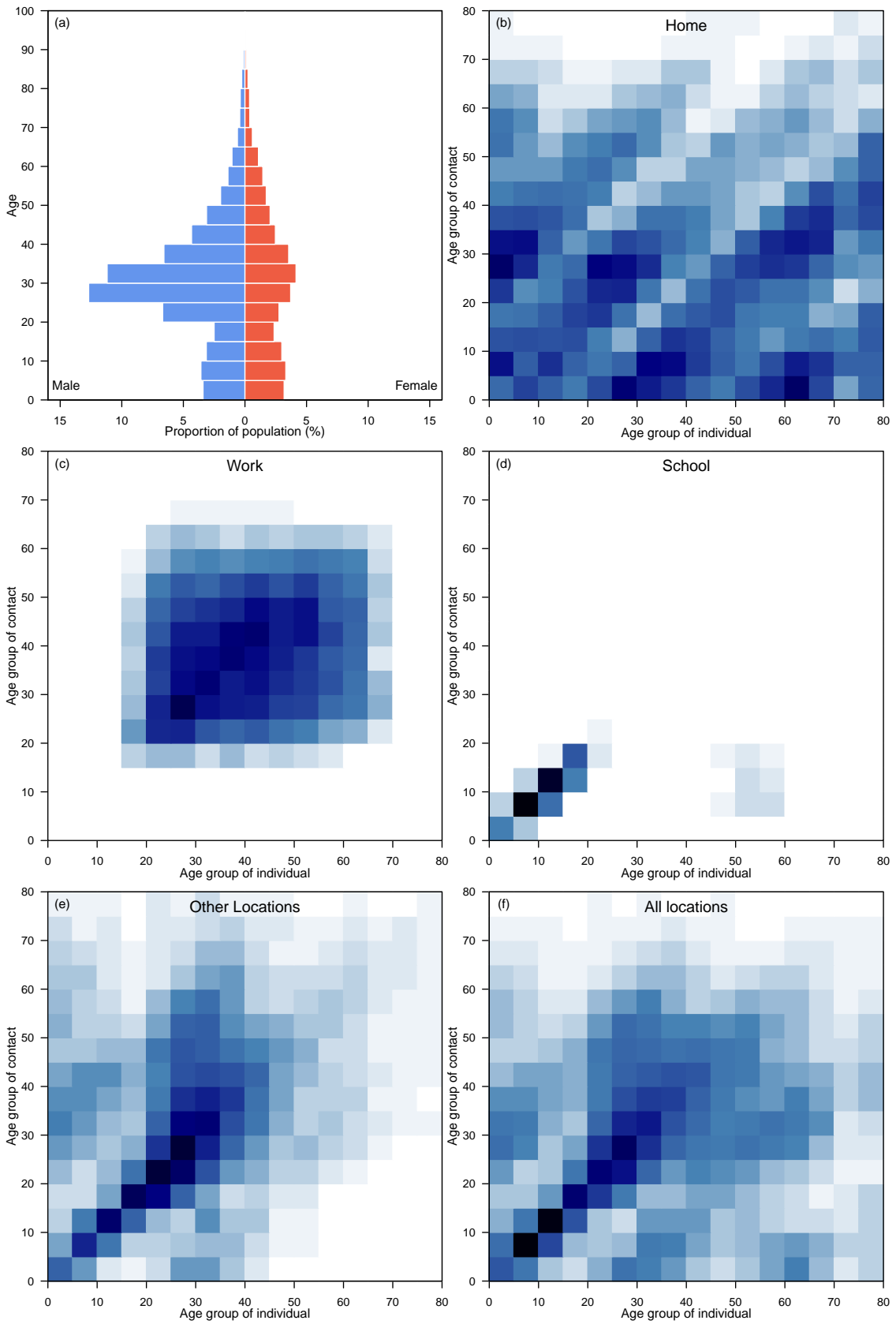
Mali
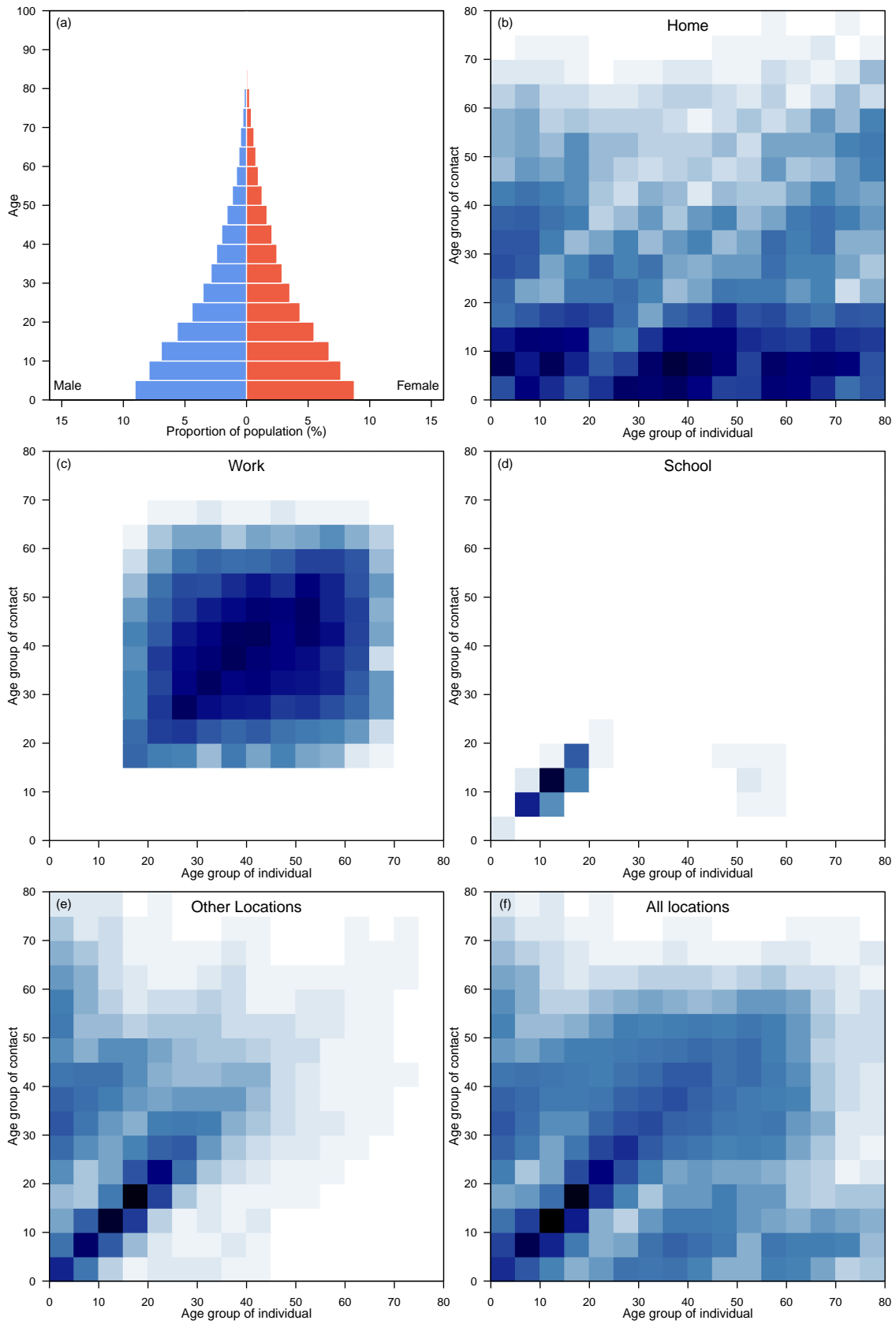
Malta
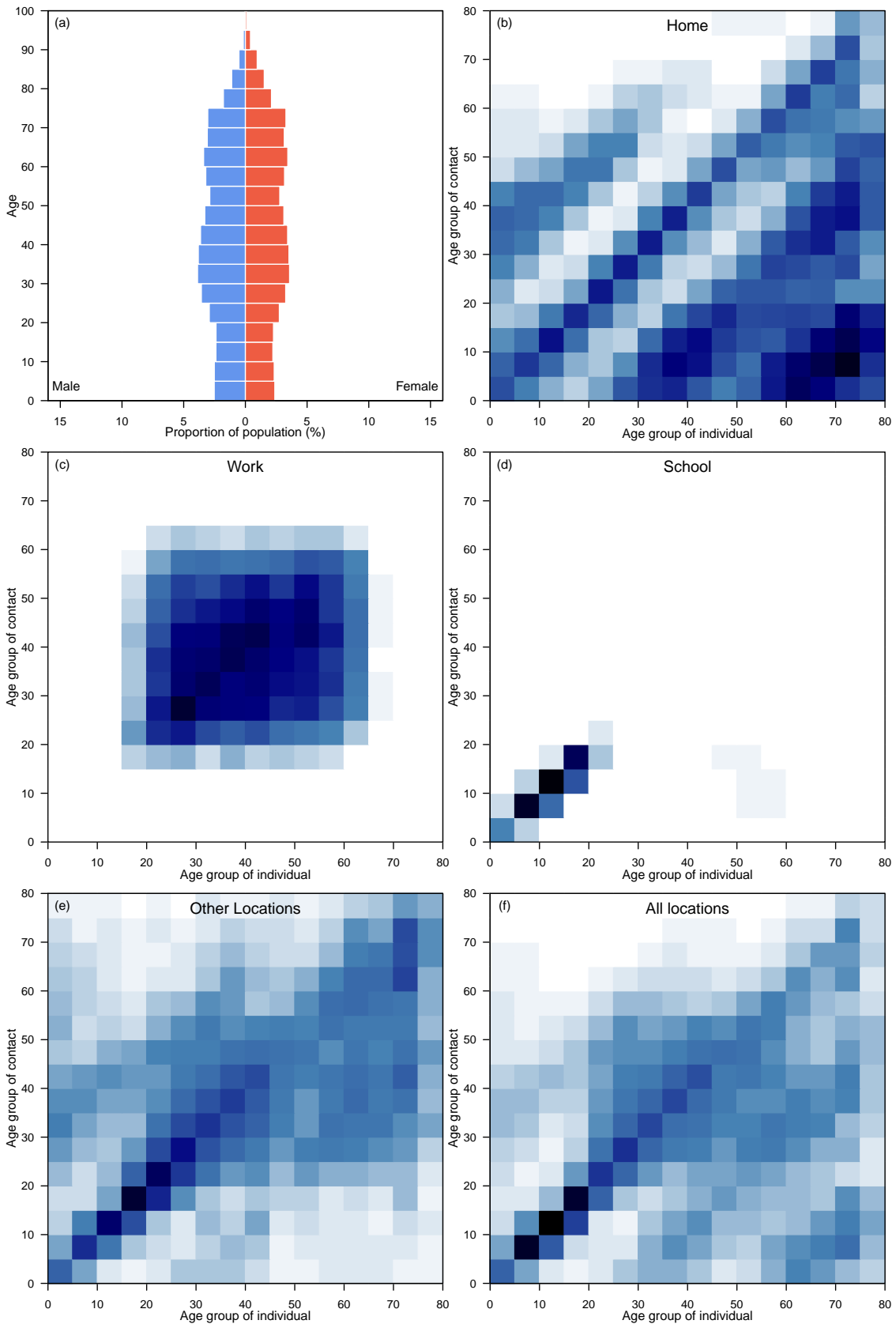
Mauritania
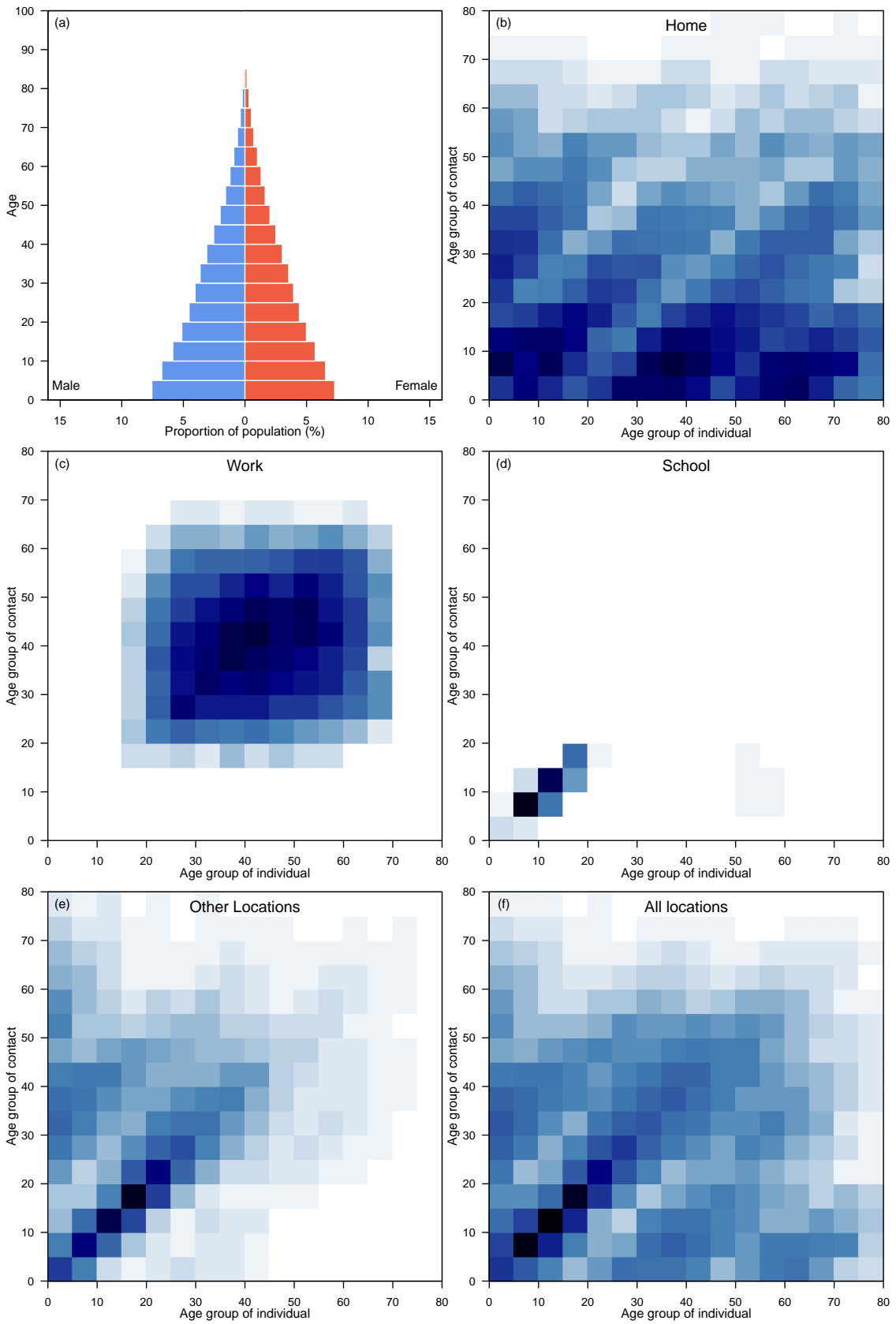
Mauritius
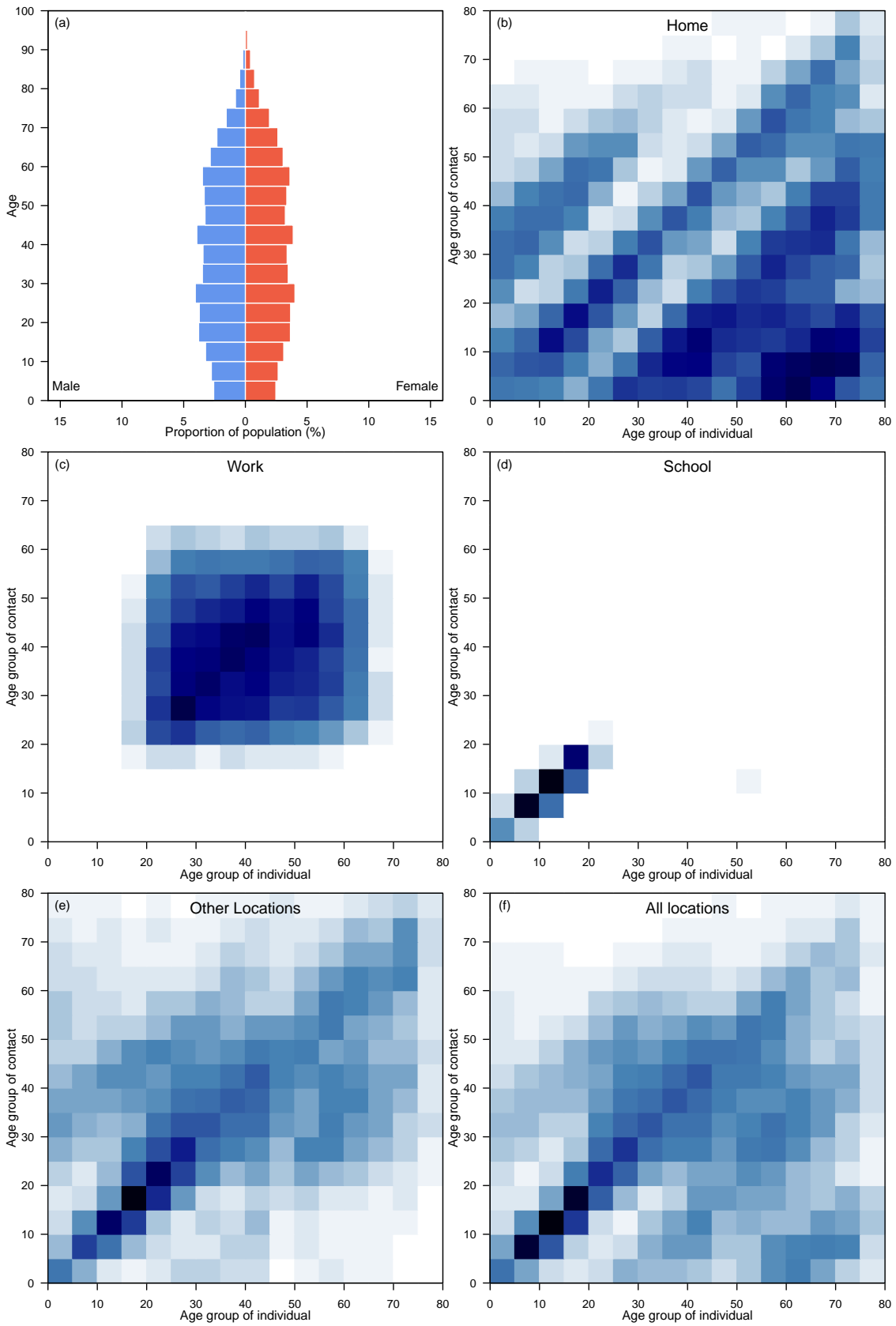
Mexico
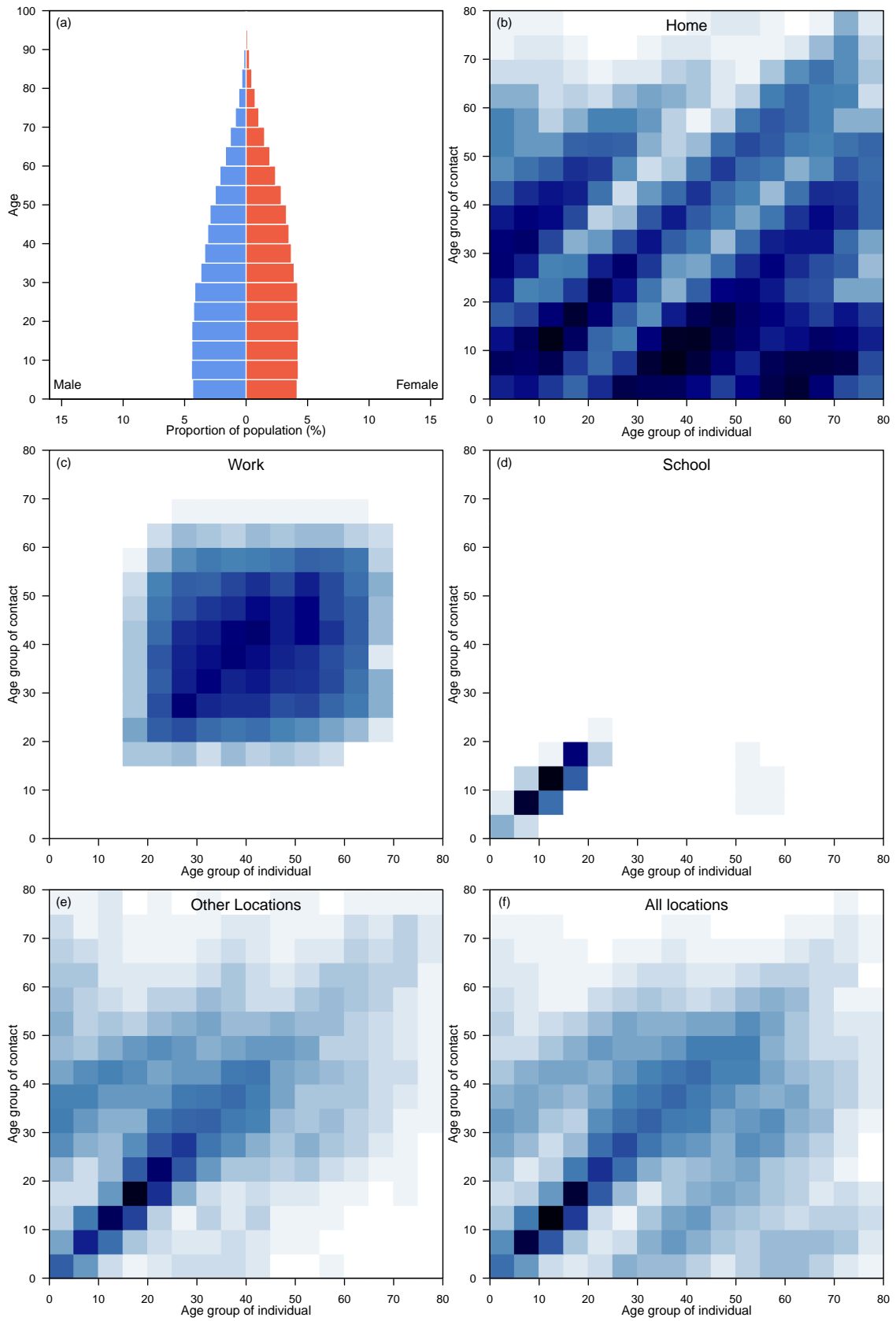
Mongolia
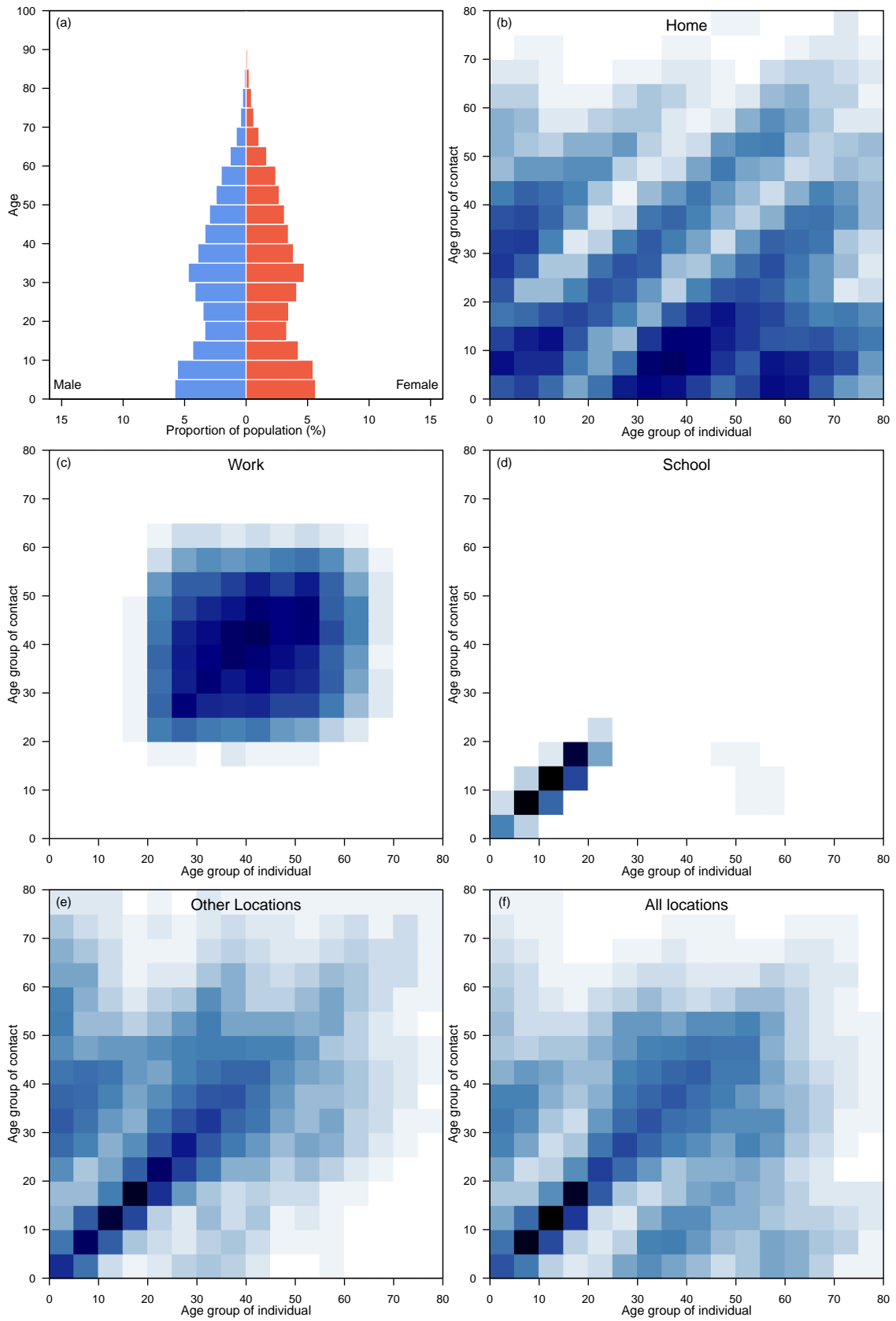
Montenegro
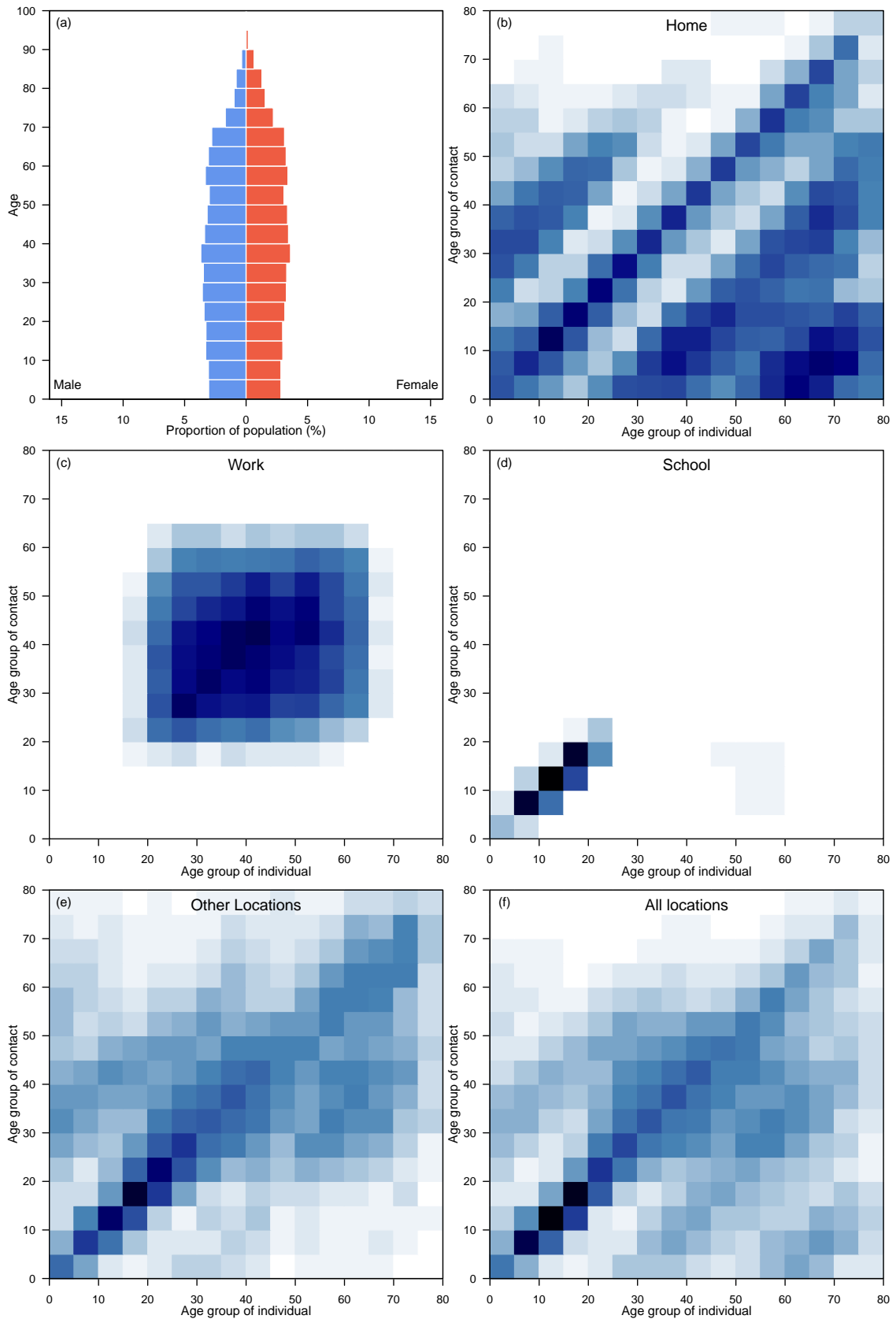
Morocco
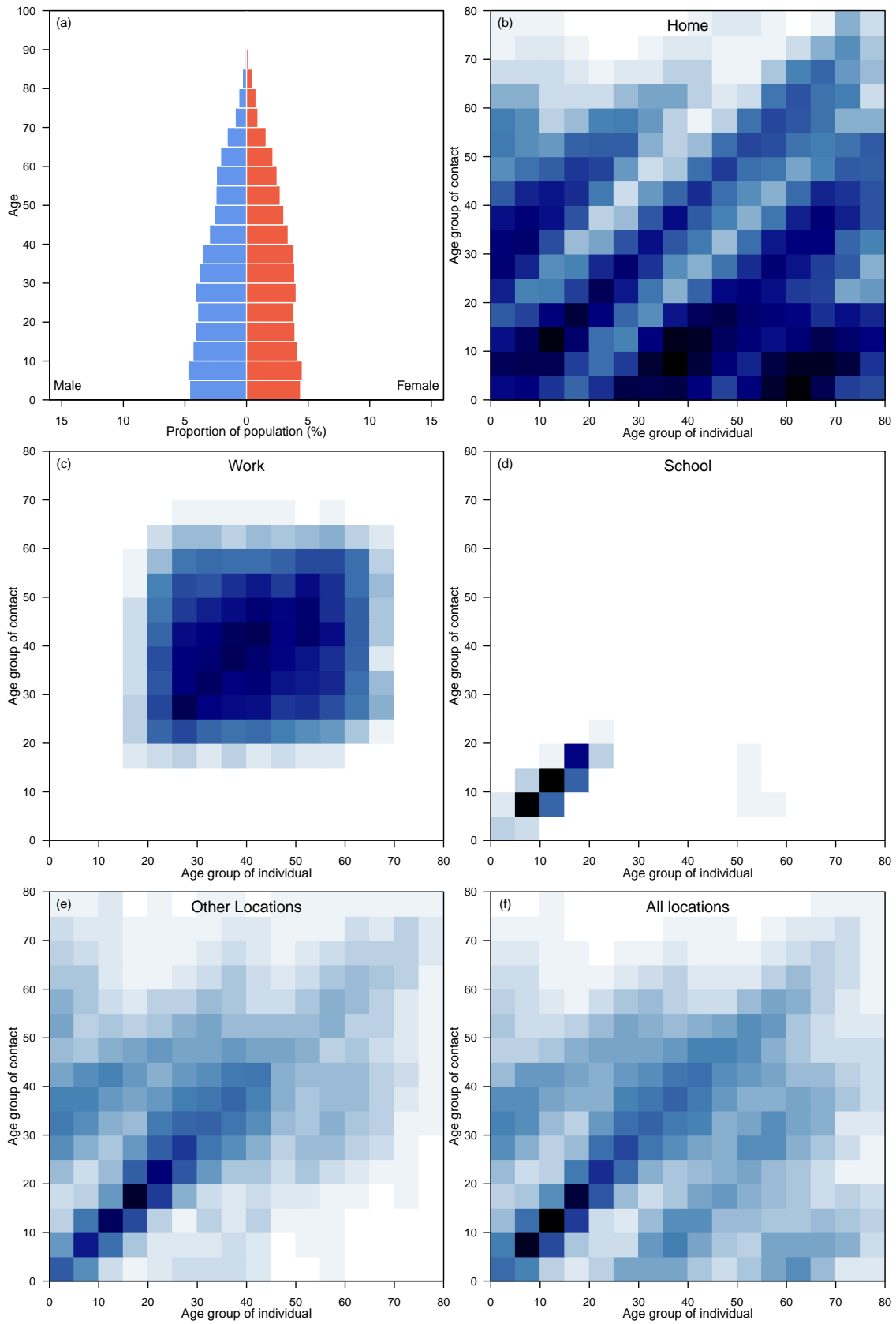
Mozambique
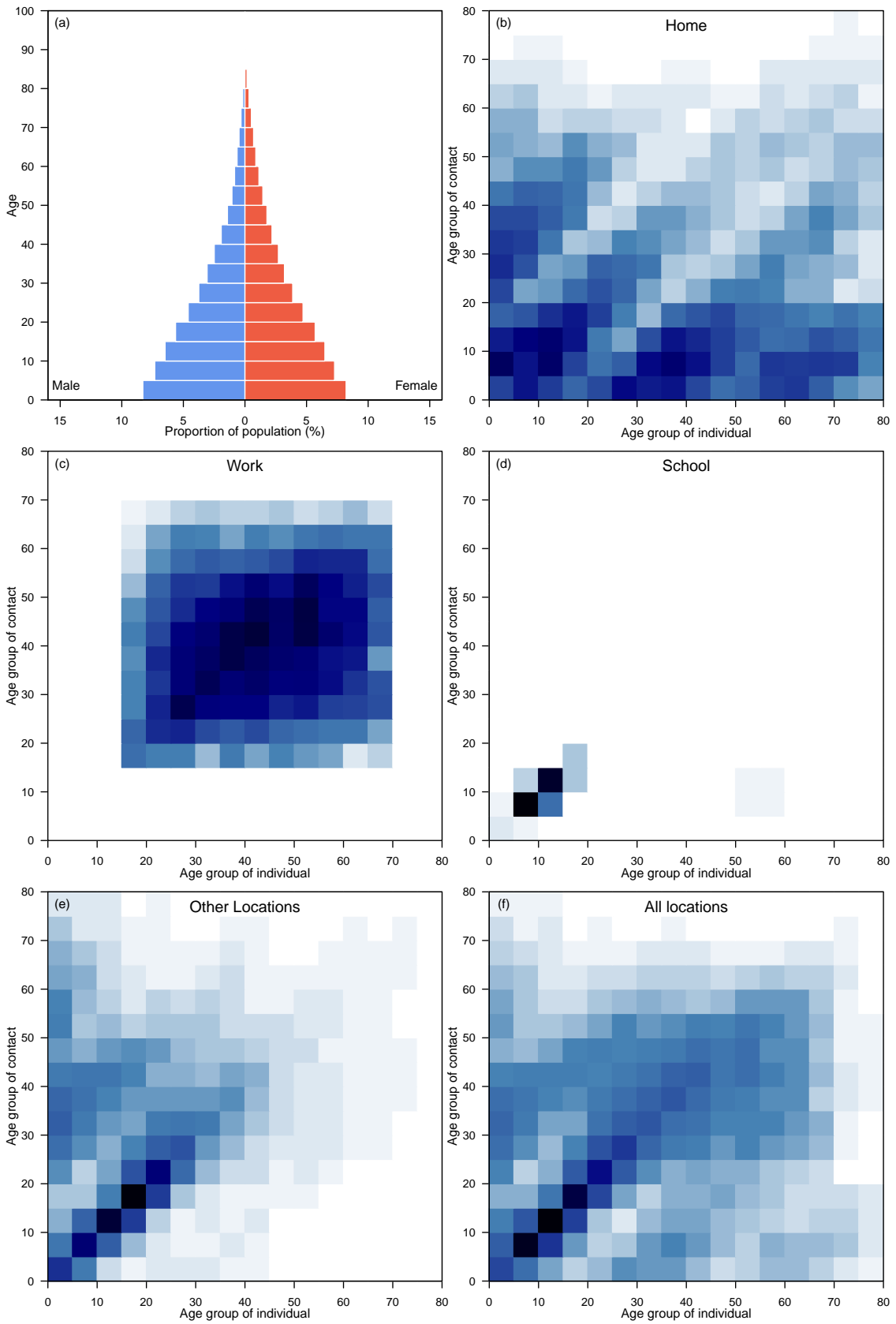
Myanmar
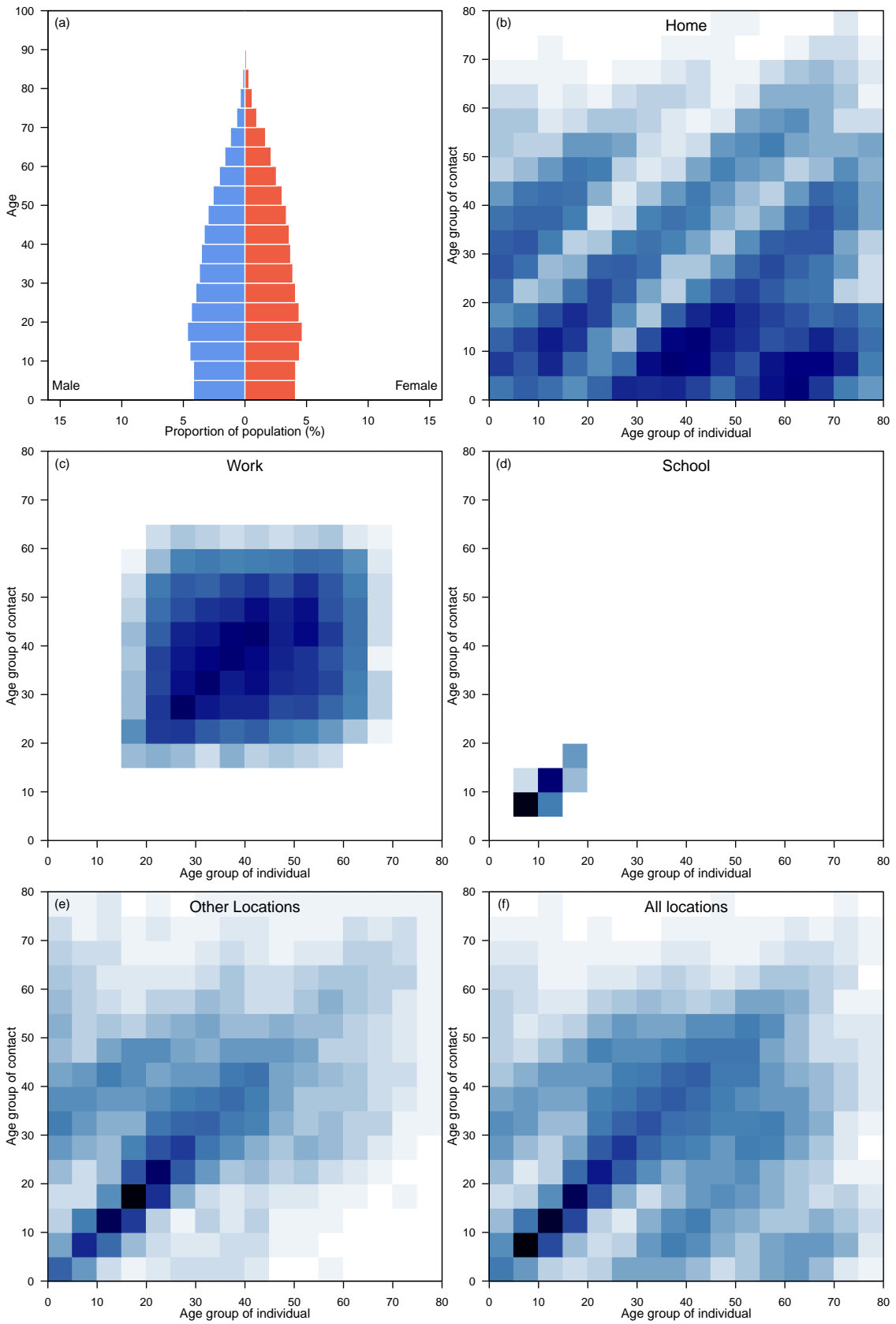
Namibia
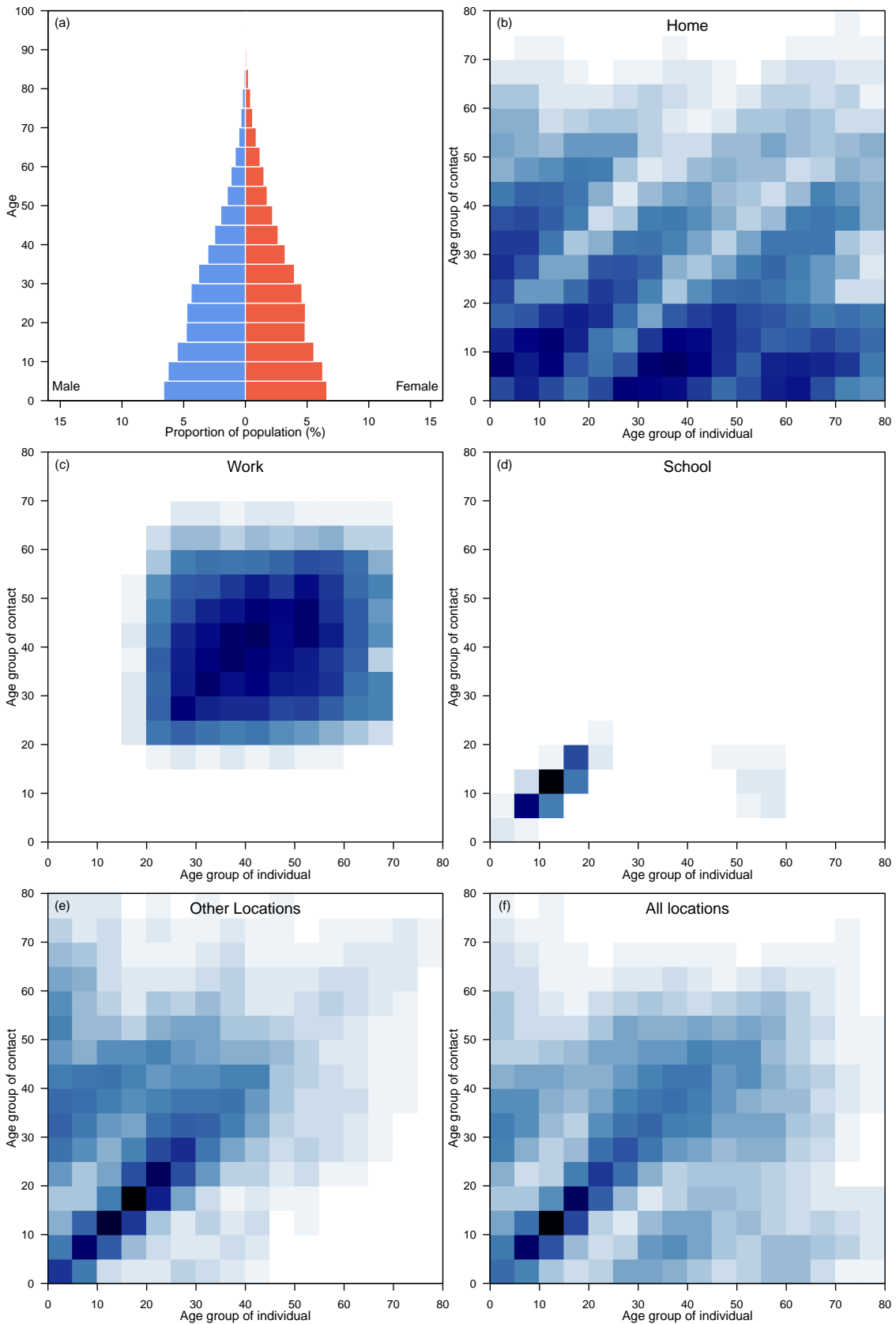
Nepal
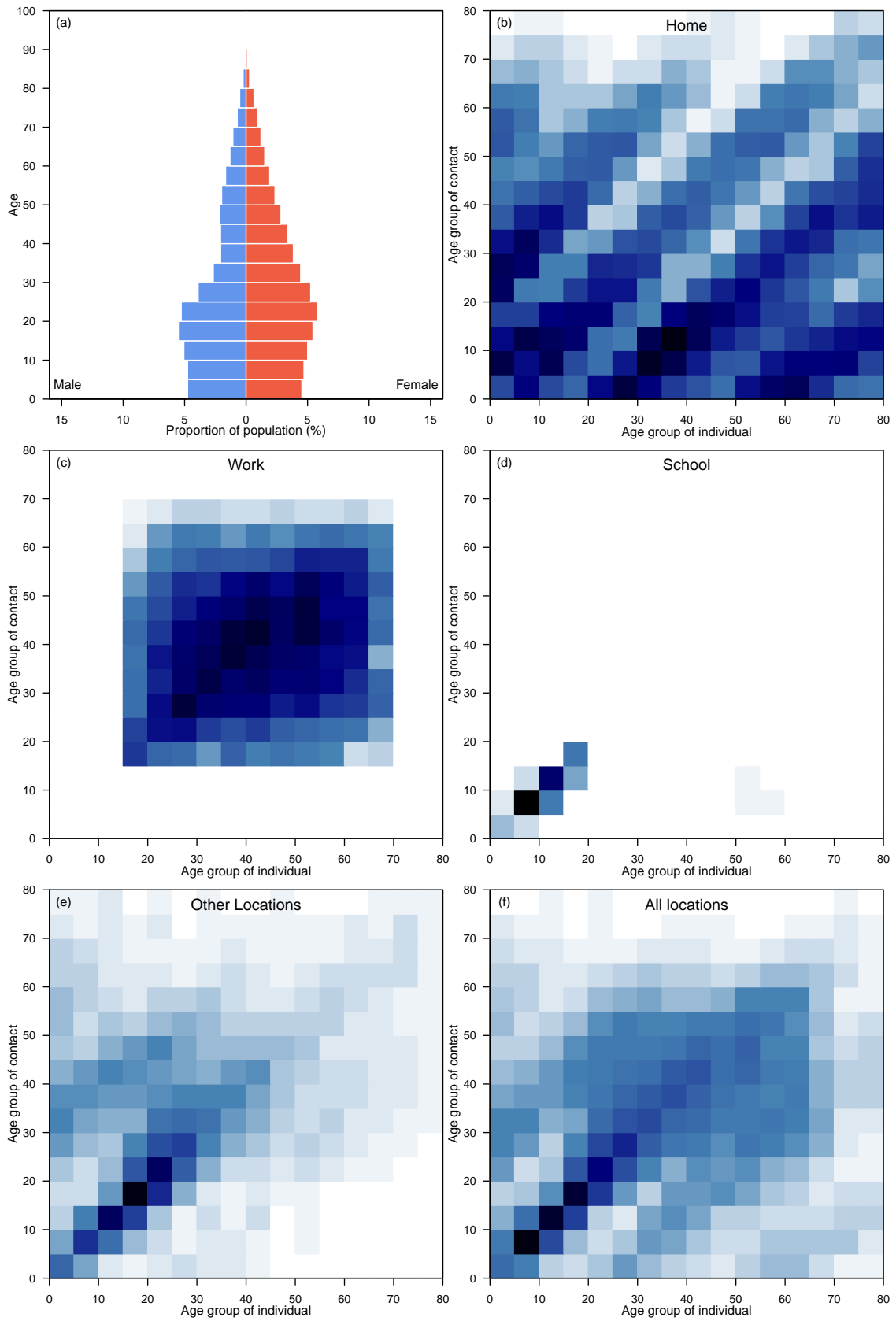
Netherlands
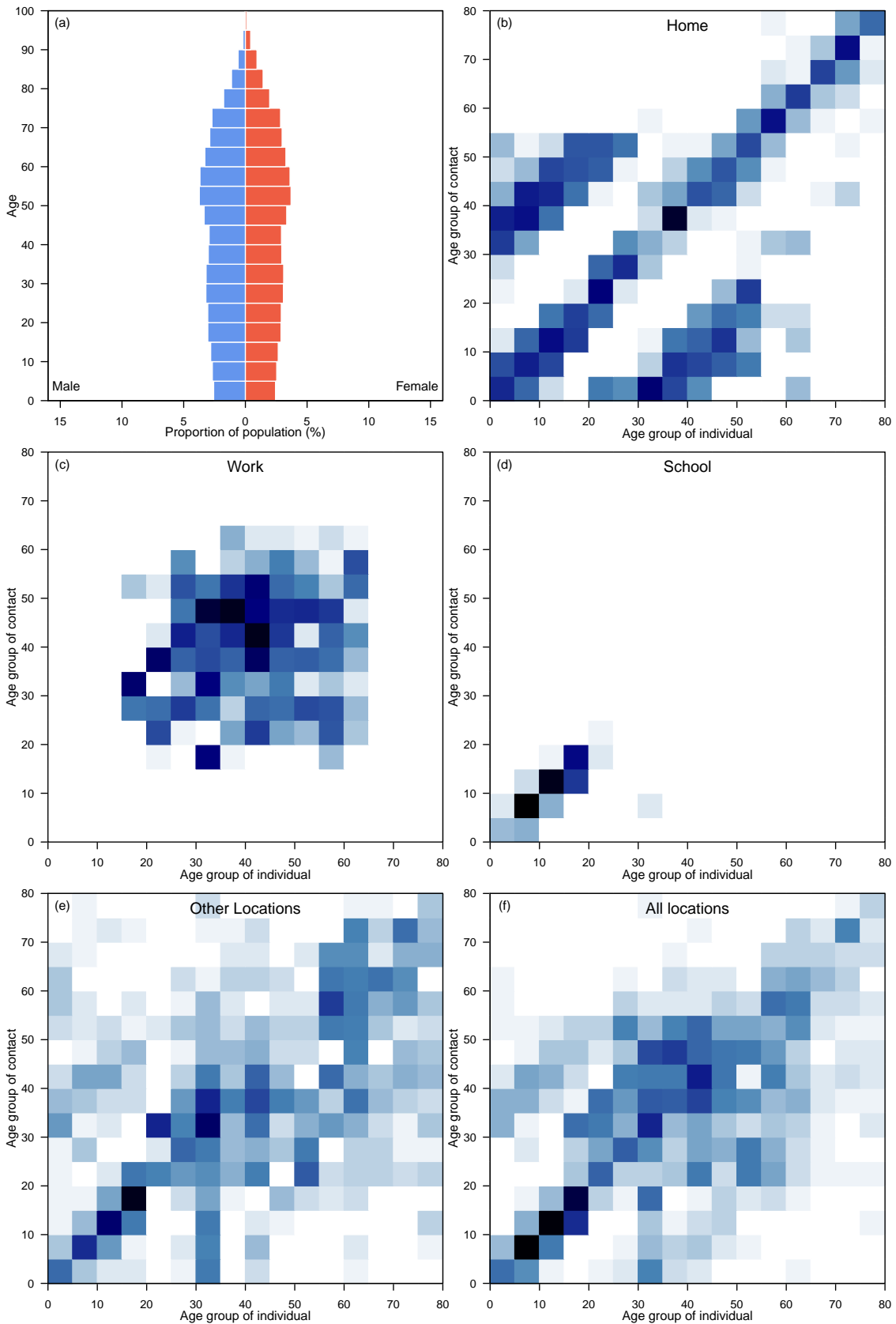
New Zealand
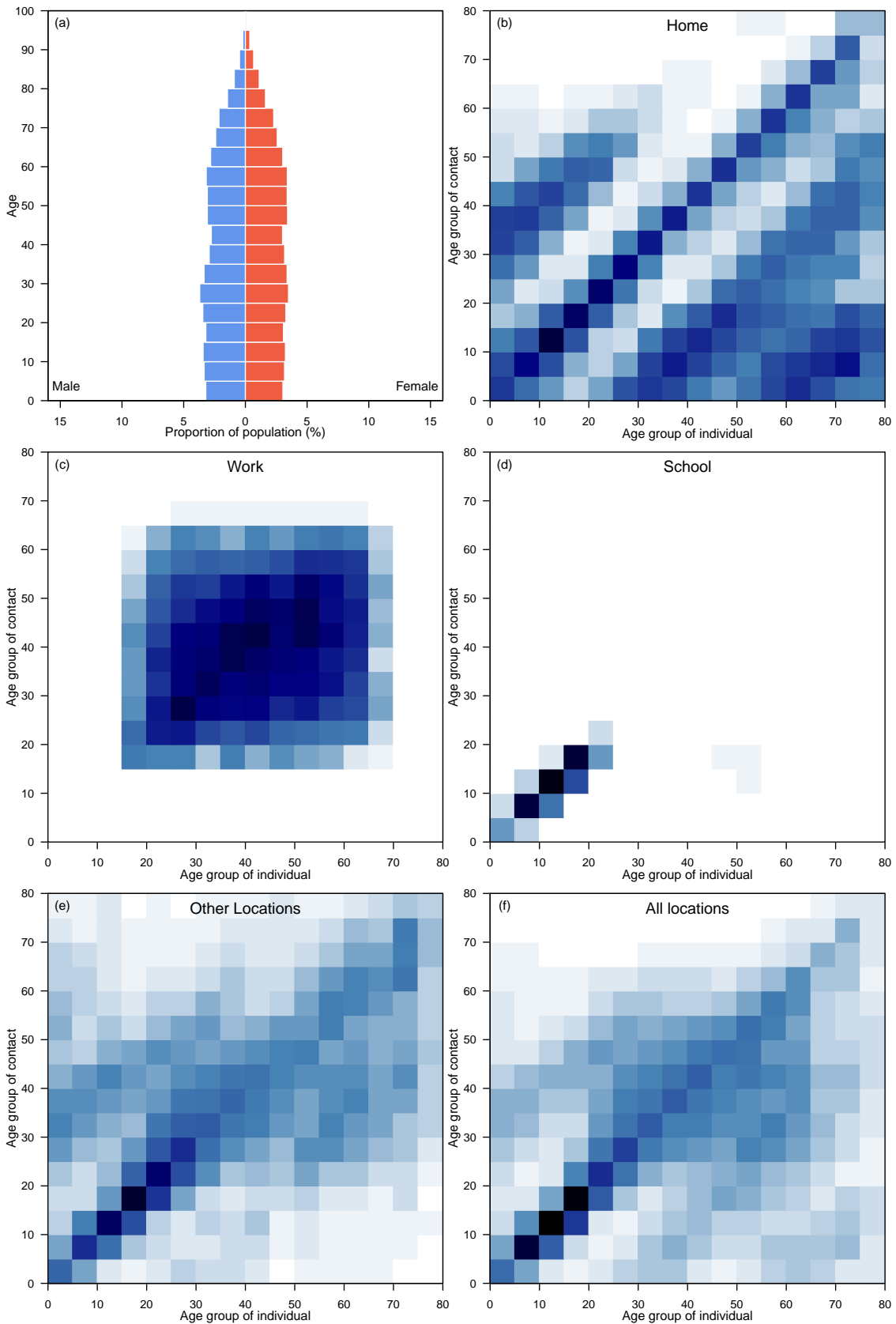
Nicaragua

Niger

Nigeria

(a)

(b) Home

(c) Work

(d) School

(e) Other Locations

(f) All locations
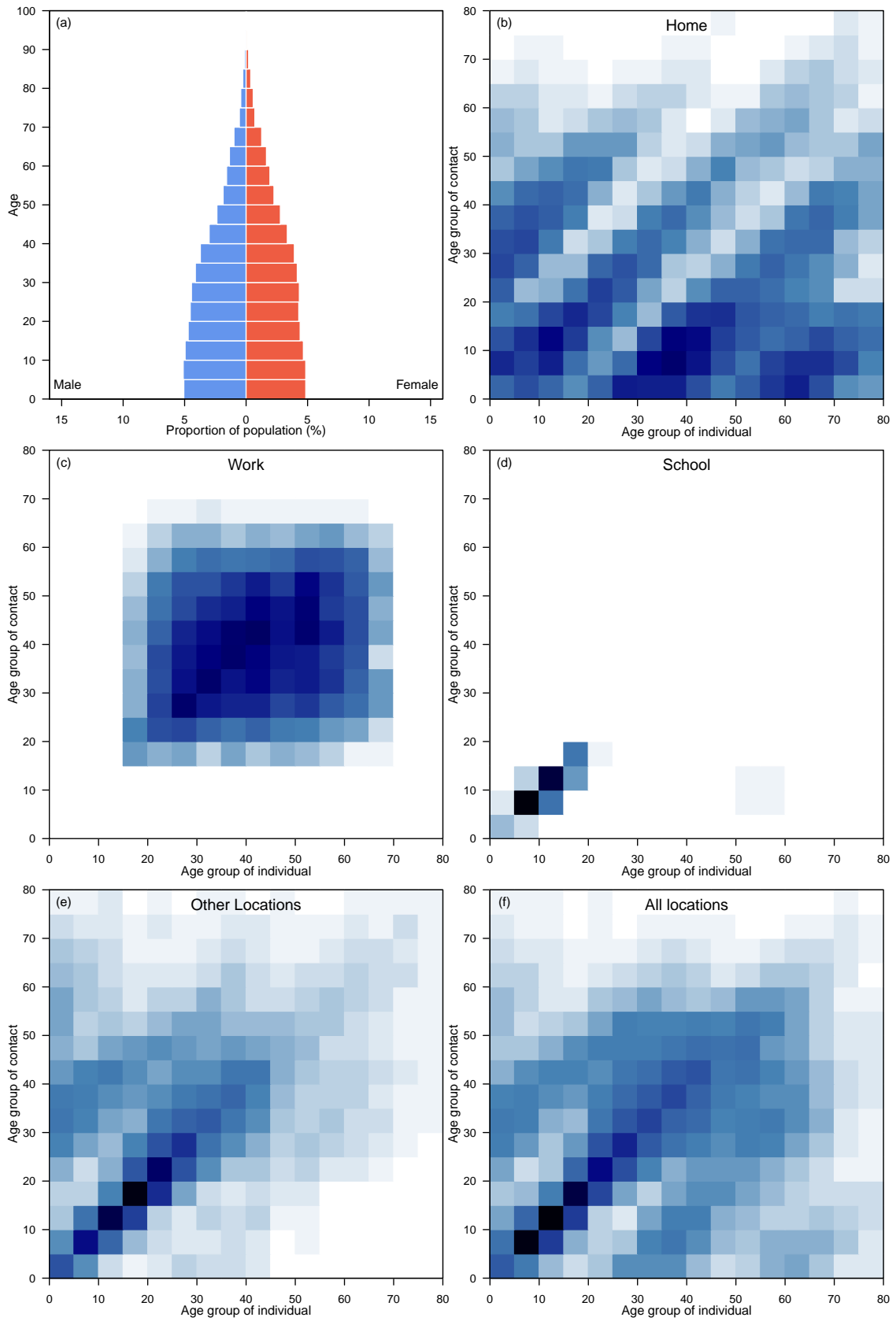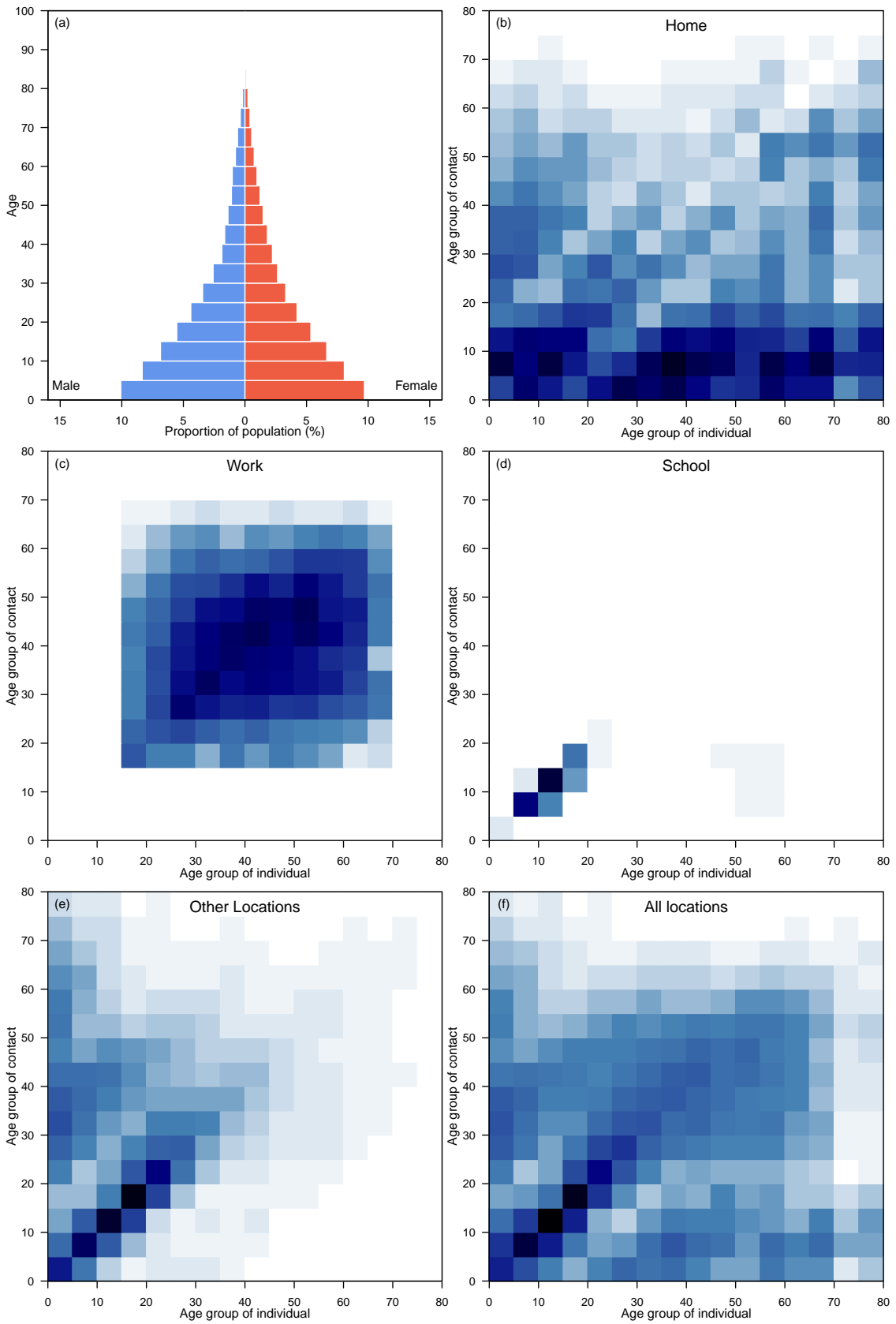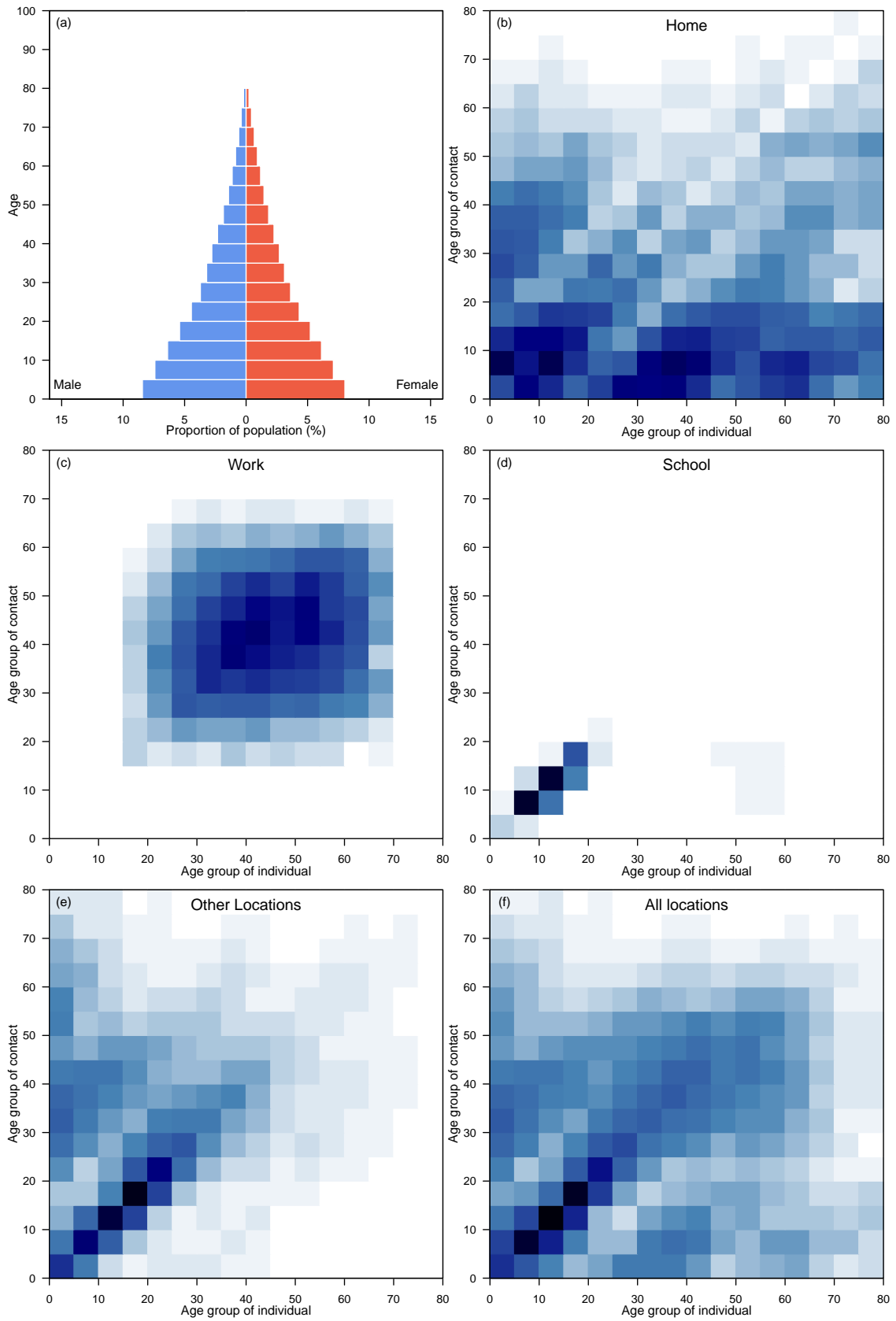
North Macedonia
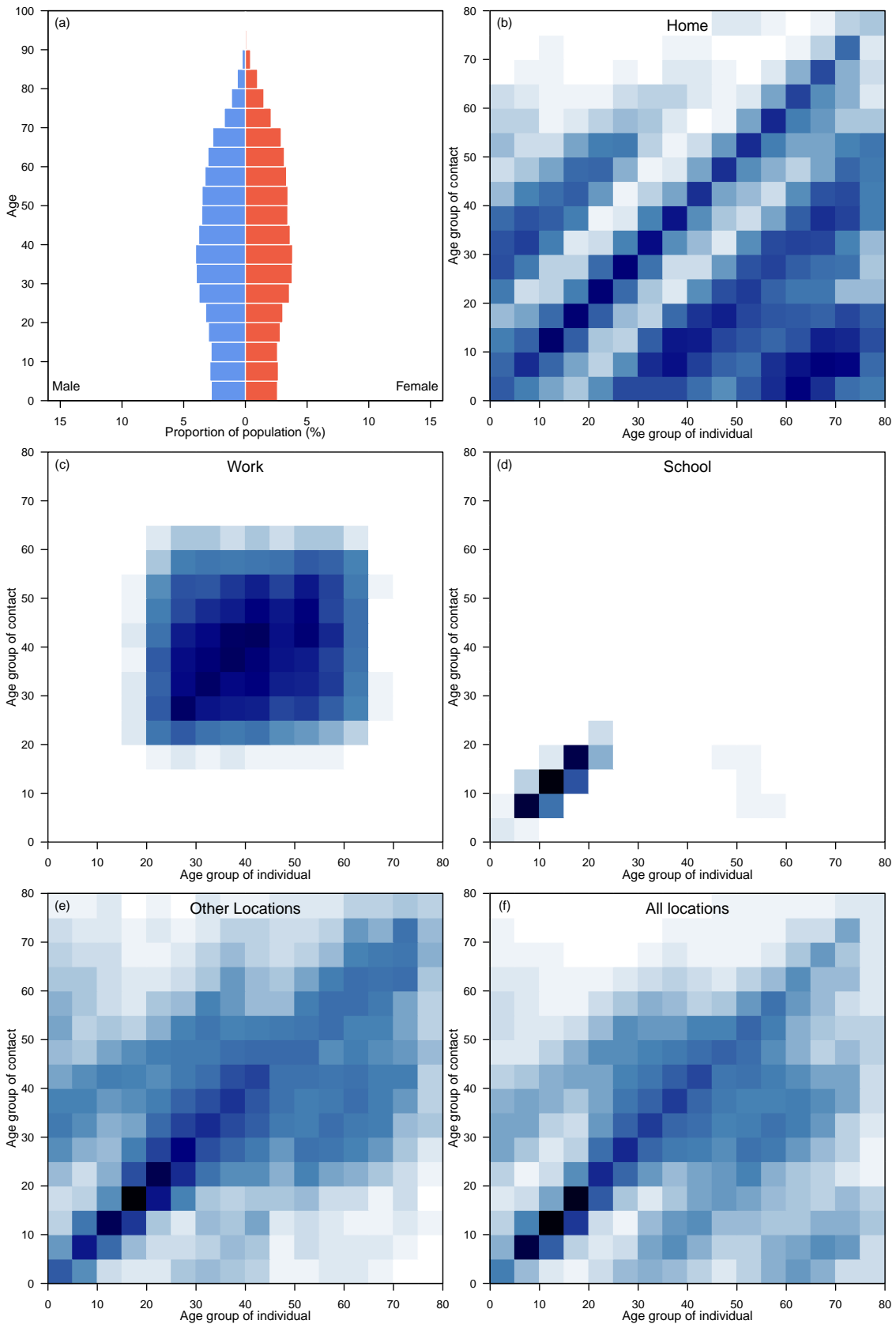
Norway
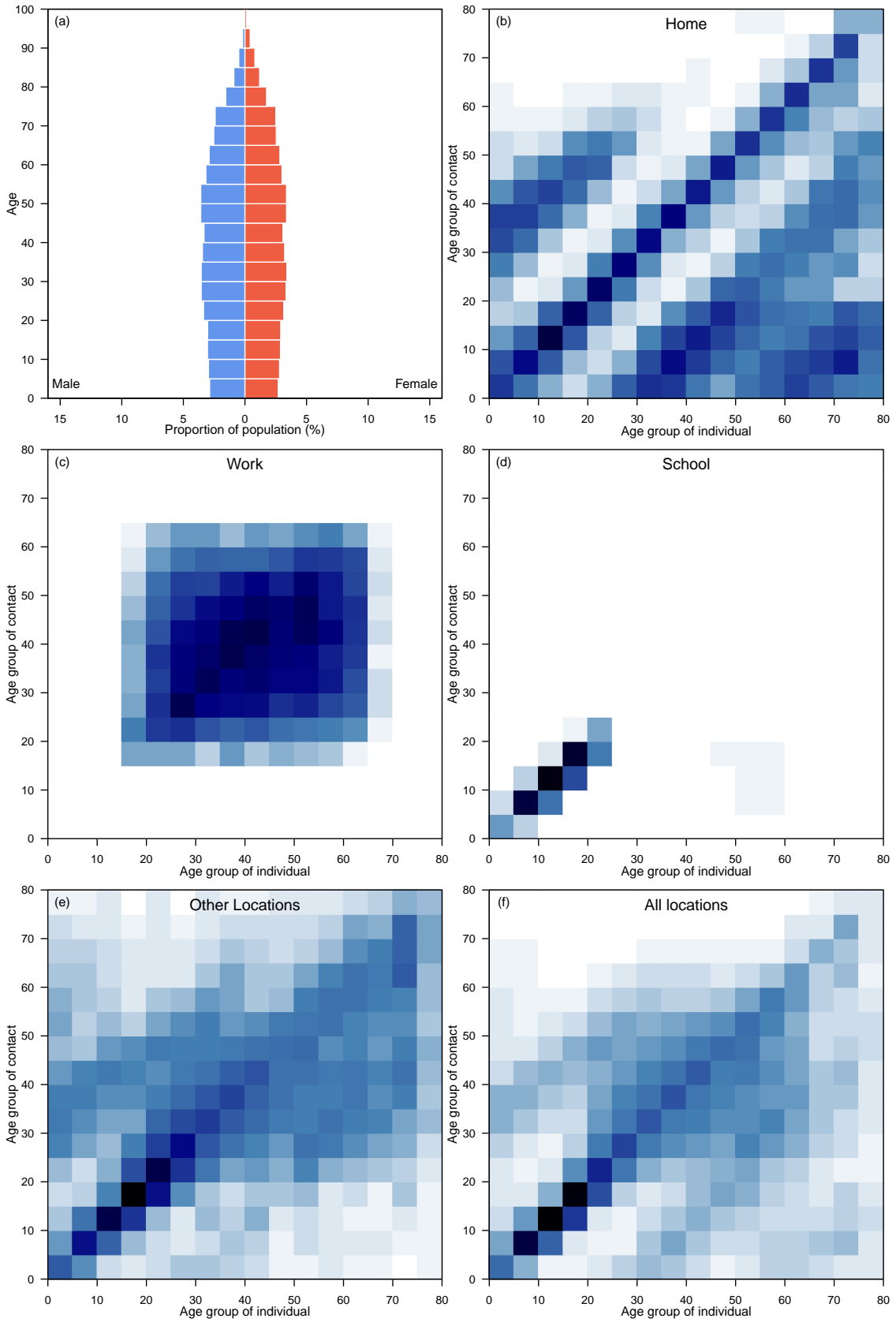
Oman
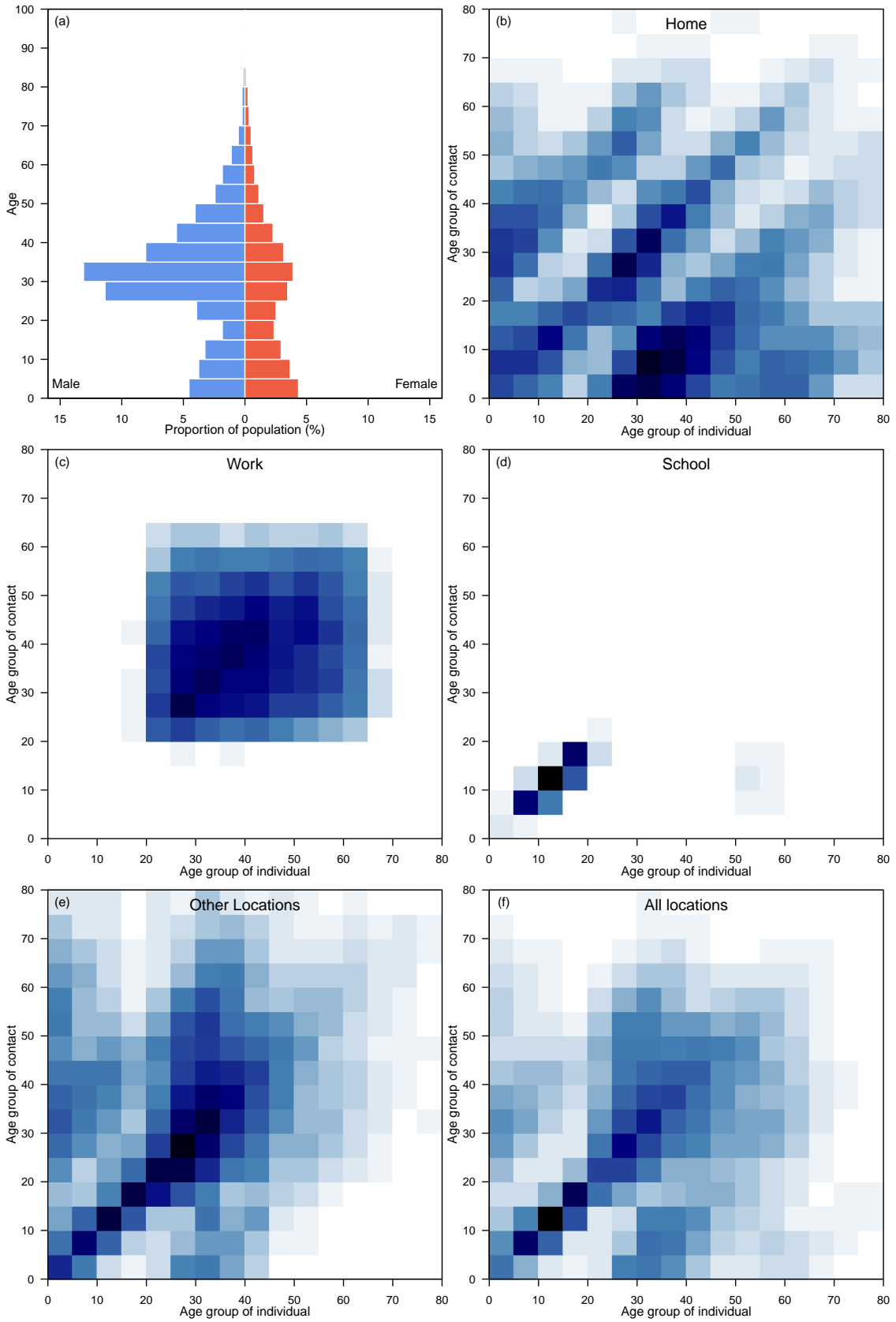
Pakistan
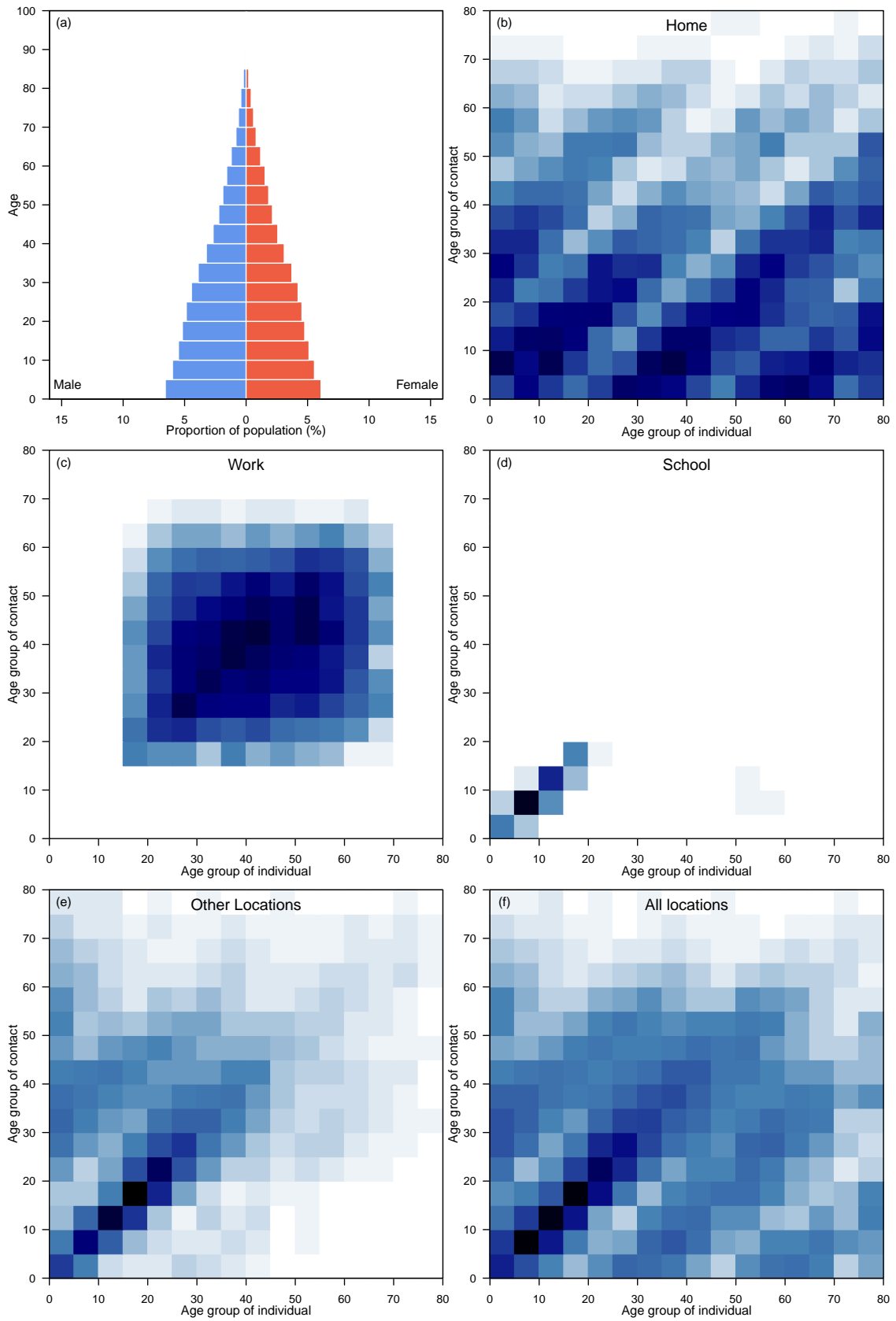
Panama
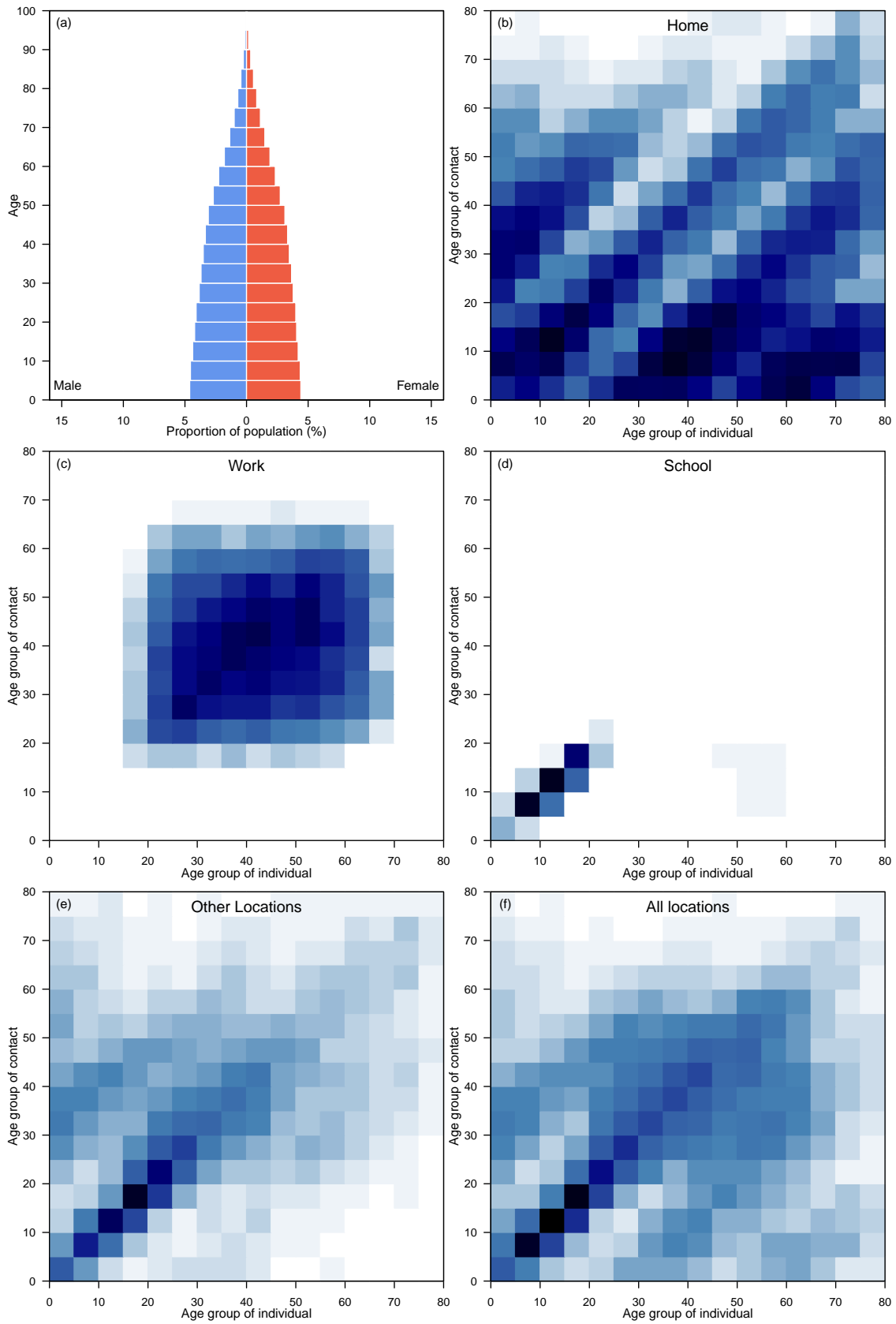
Papua New Guinea
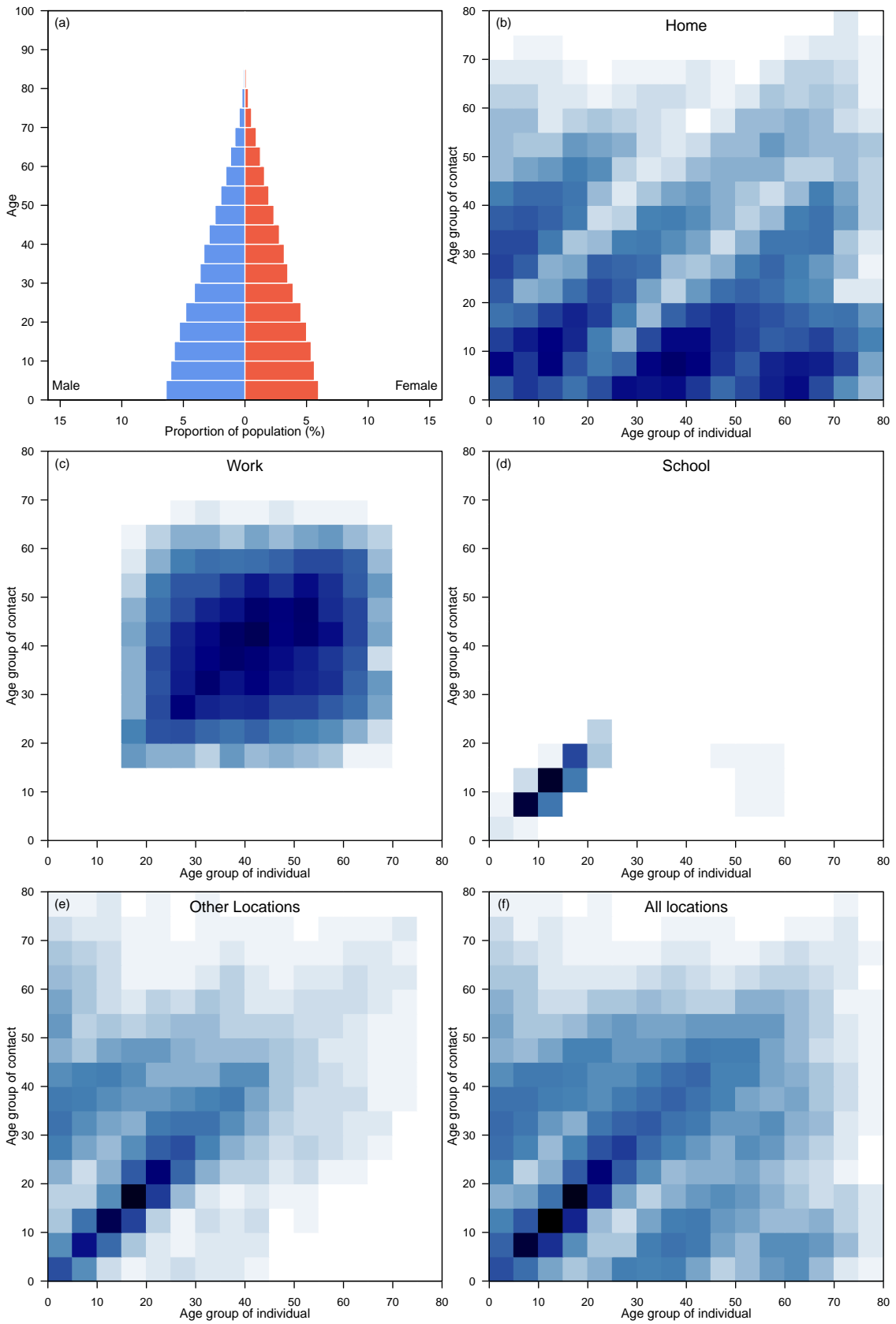
Paraguay
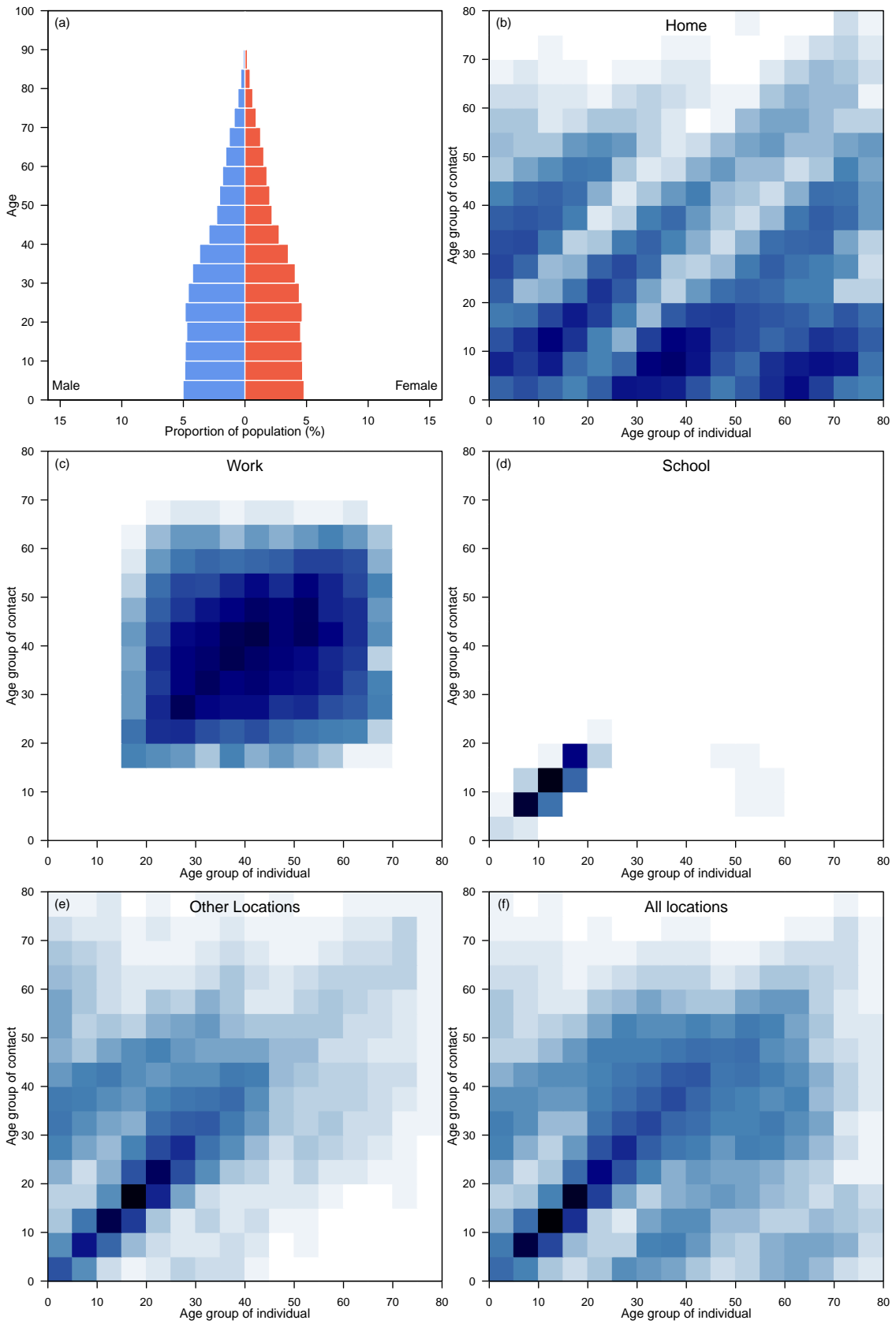
Peru
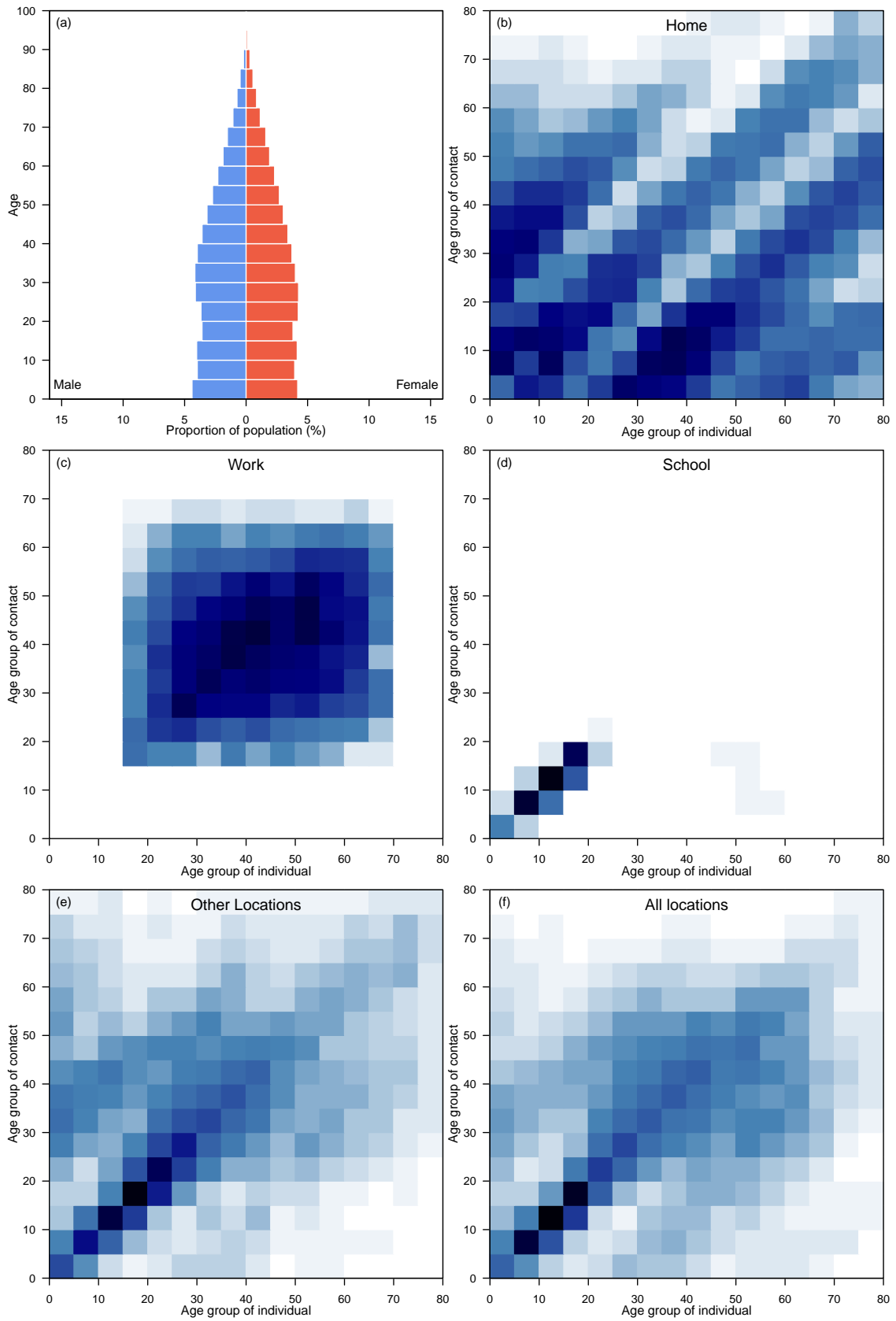
Philippines
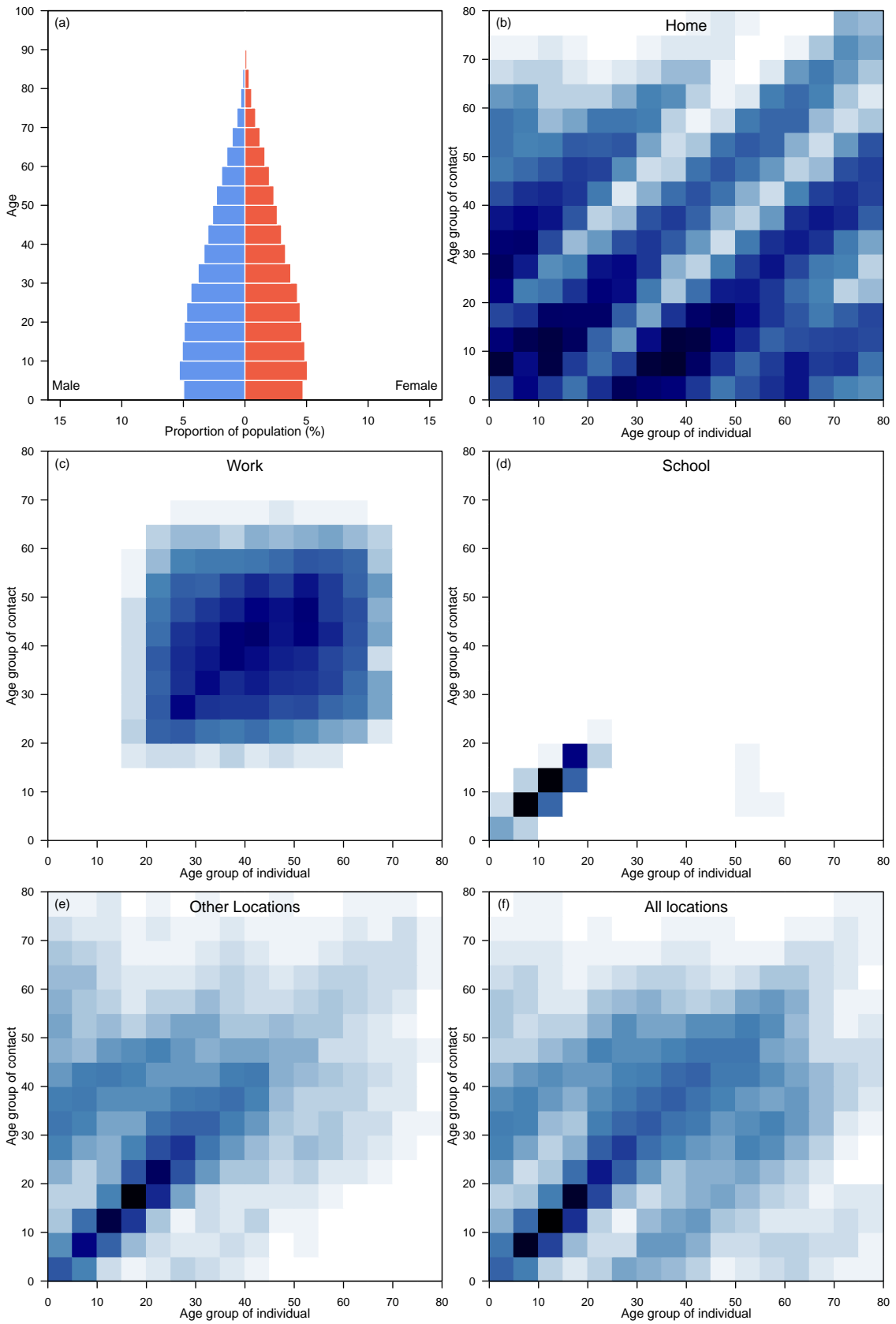
Poland
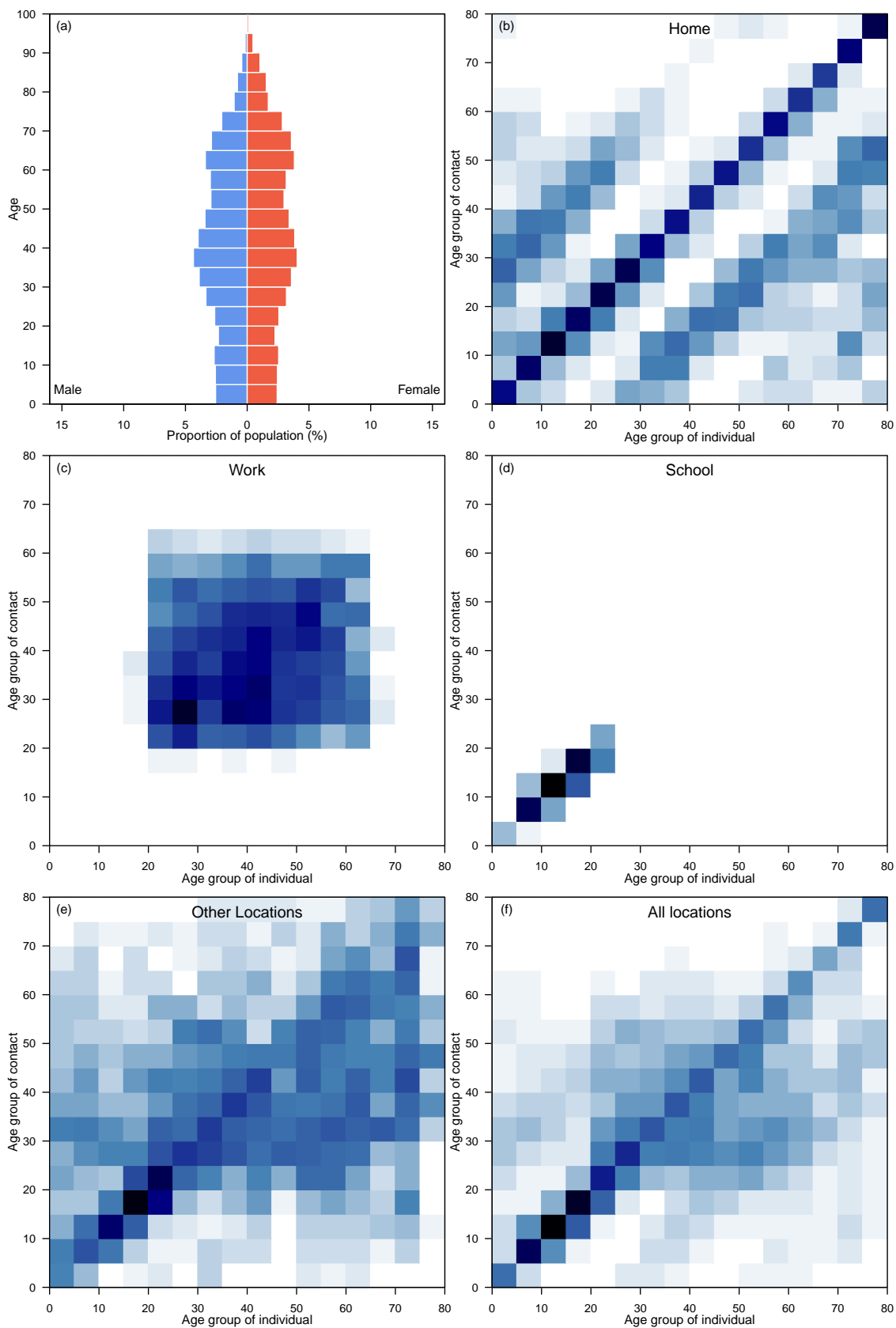
Portugal
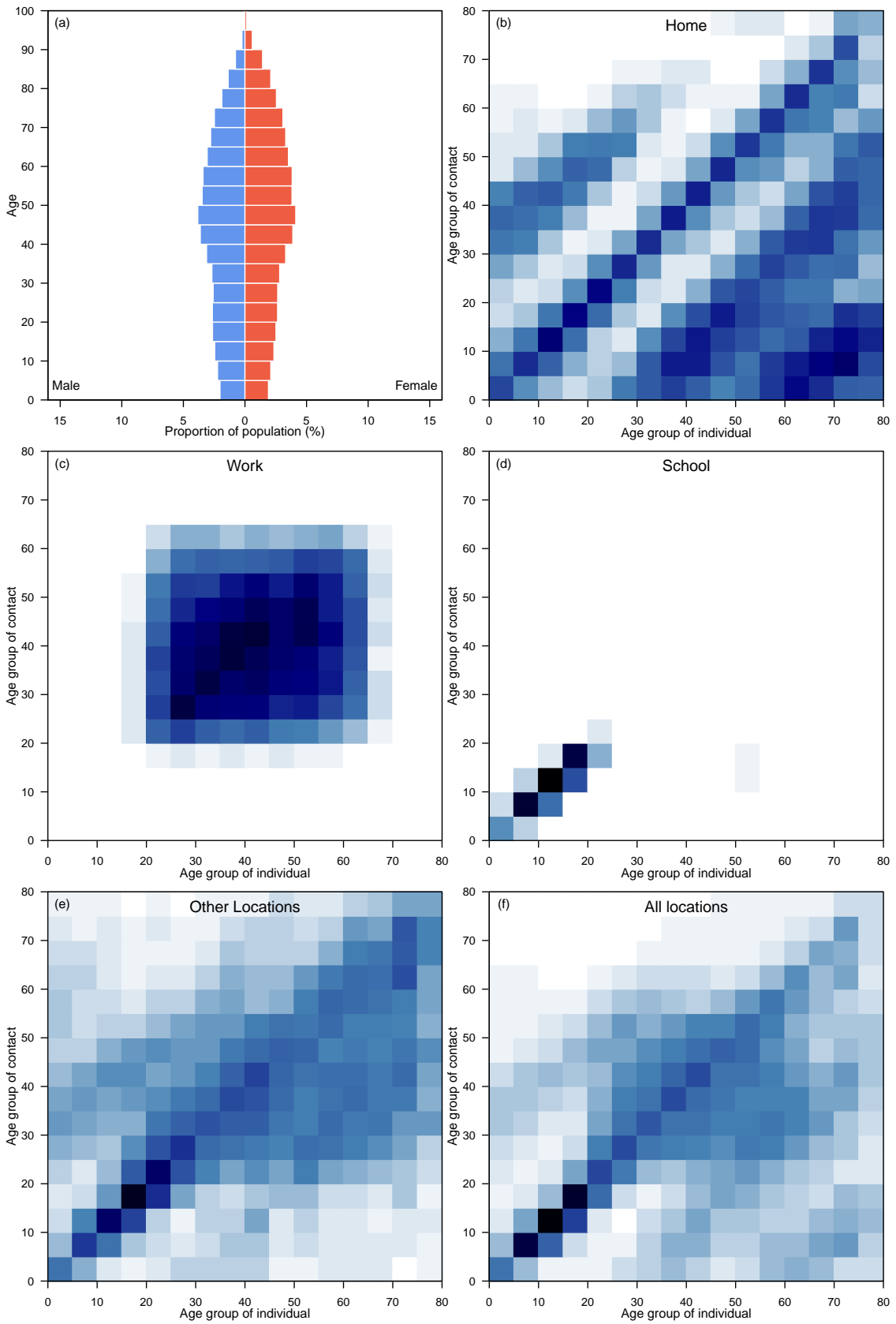
Puerto Rico
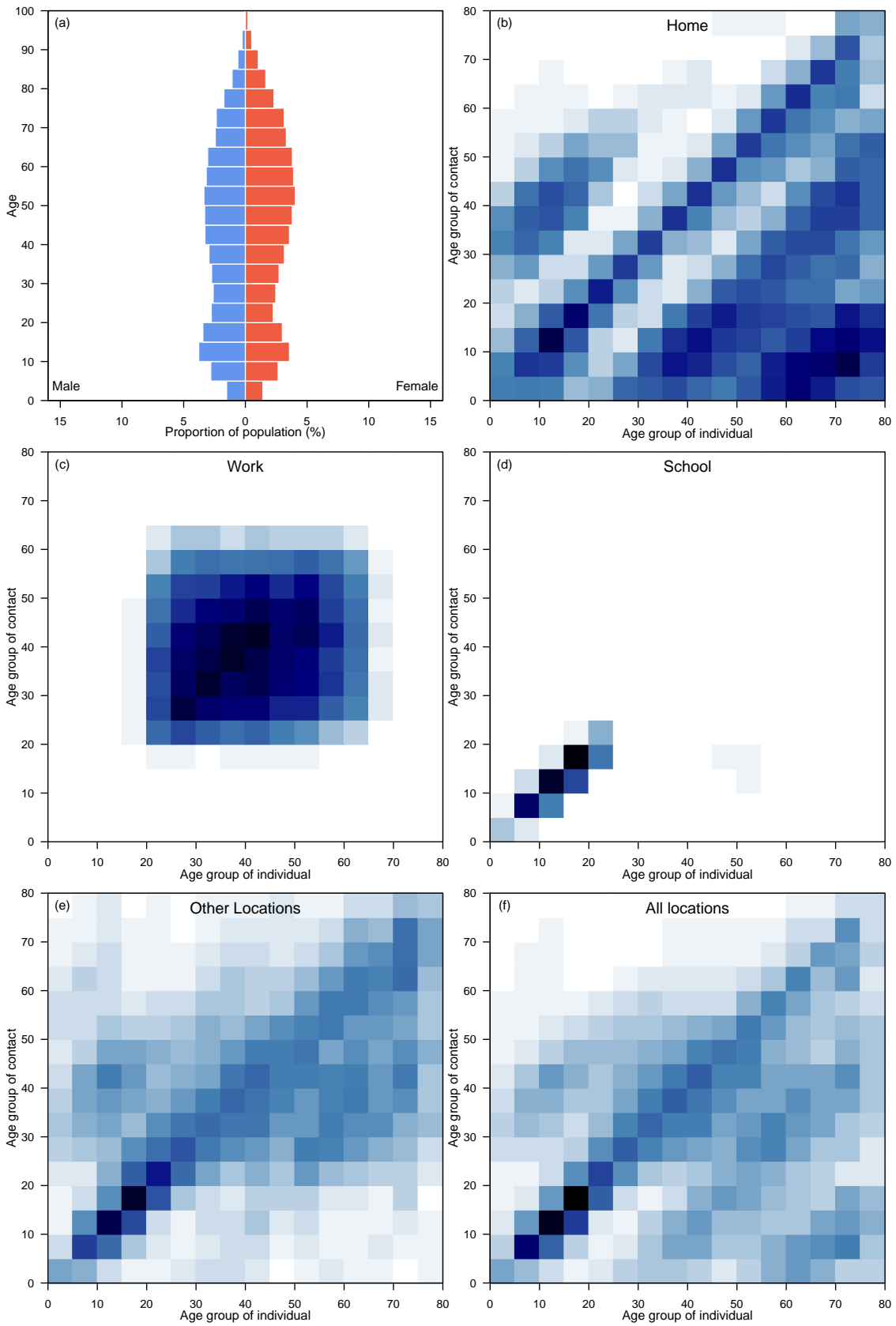
Qatar
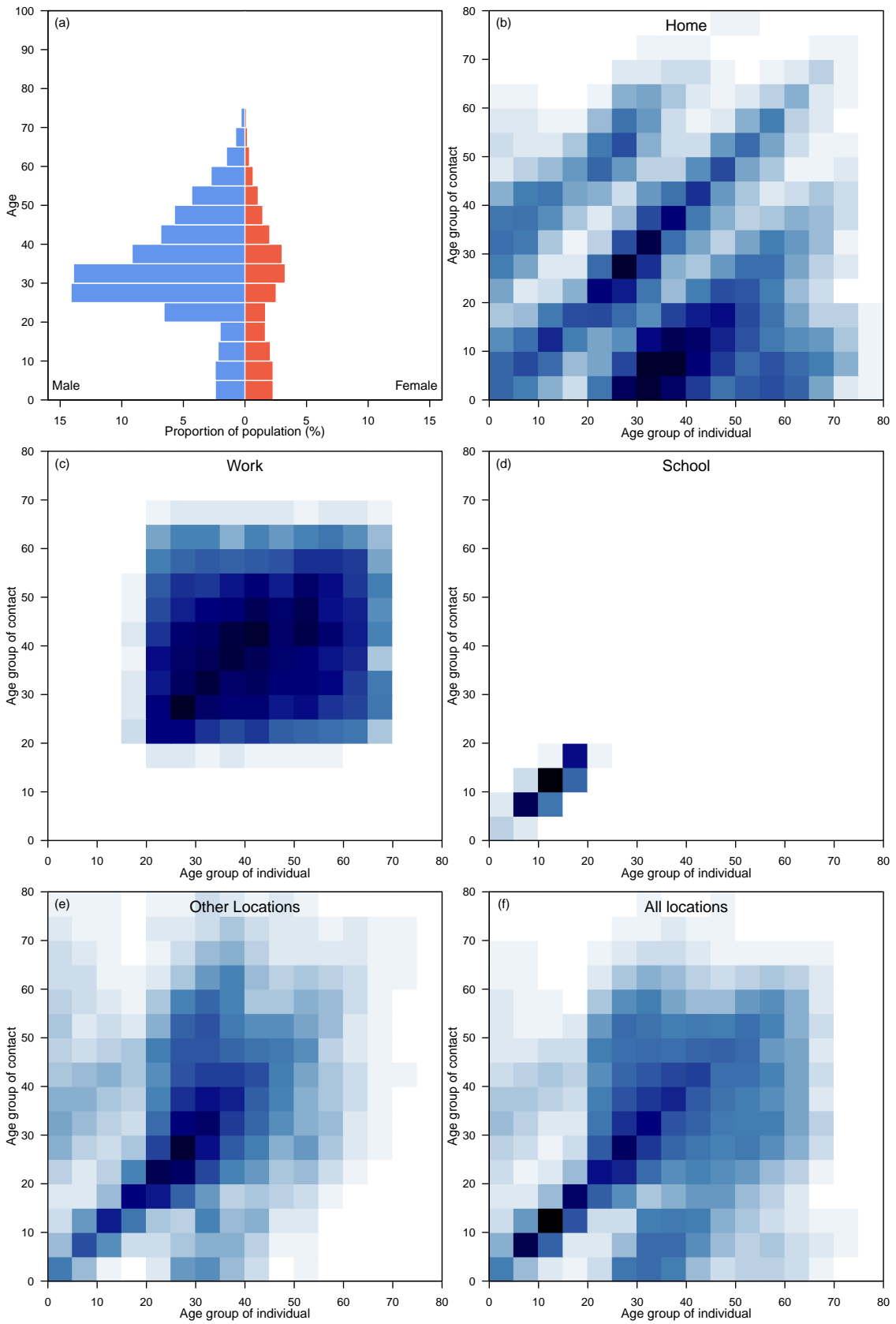
Republic of Korea
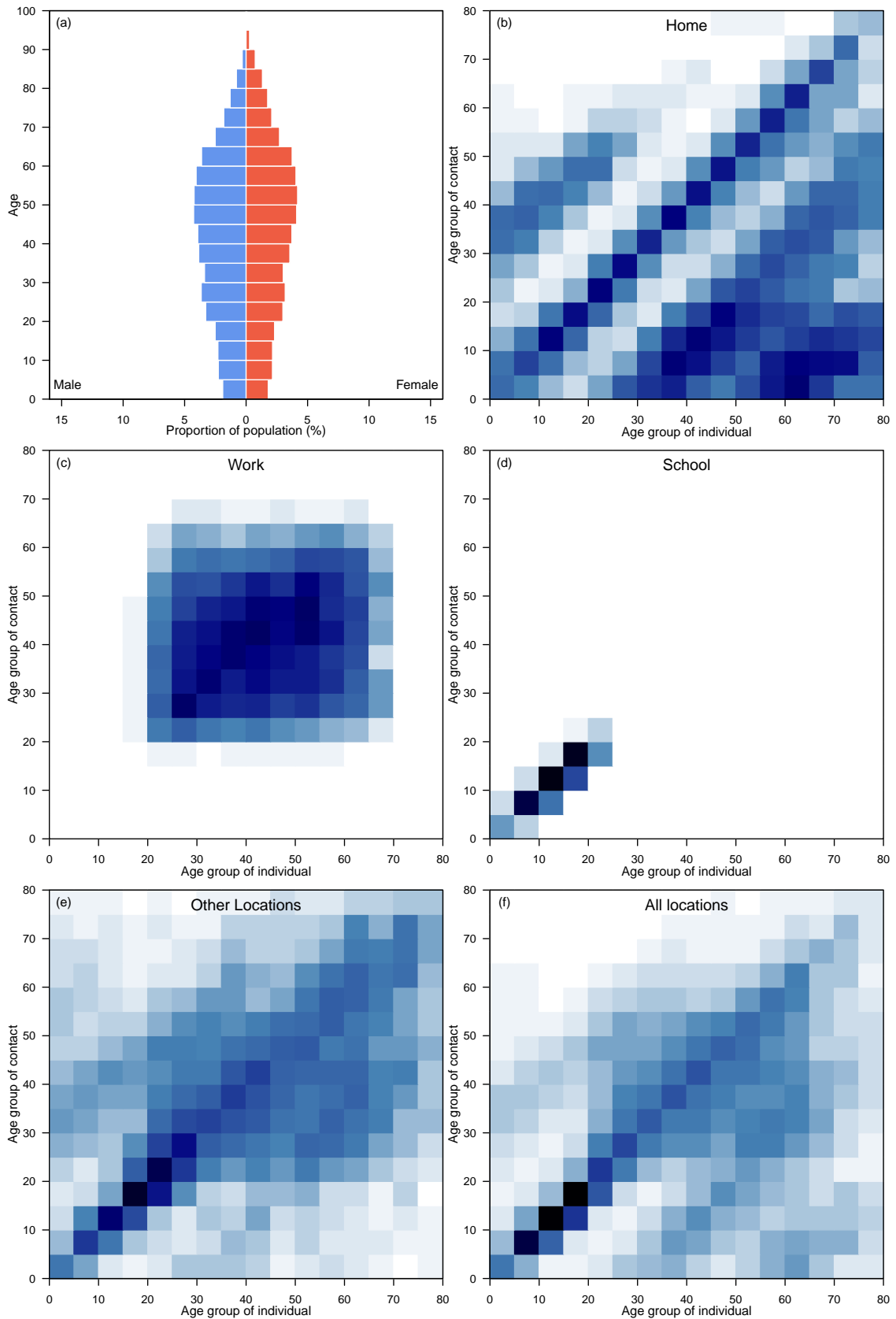
Republic of Moldova
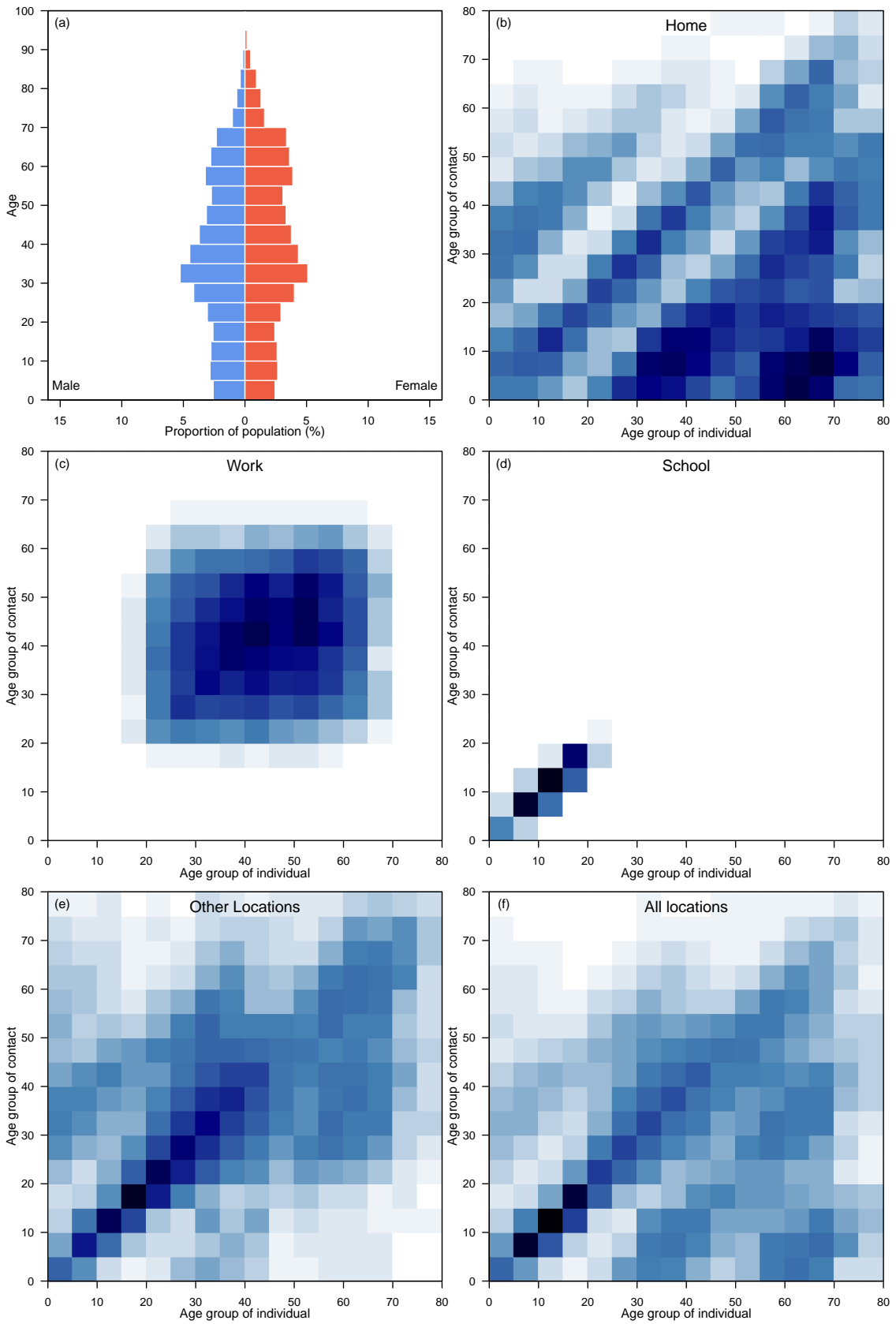
Romania
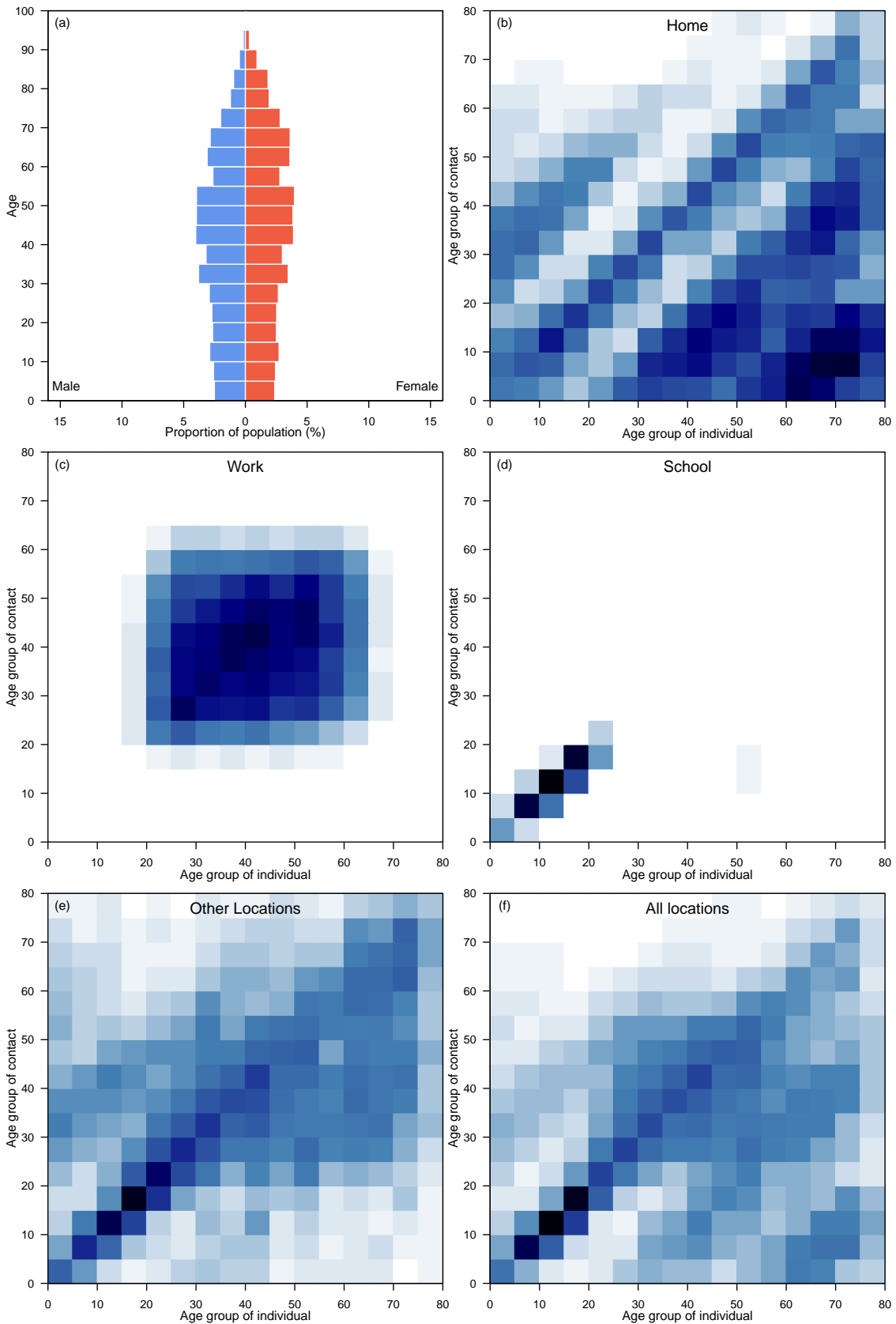
Russian Federation
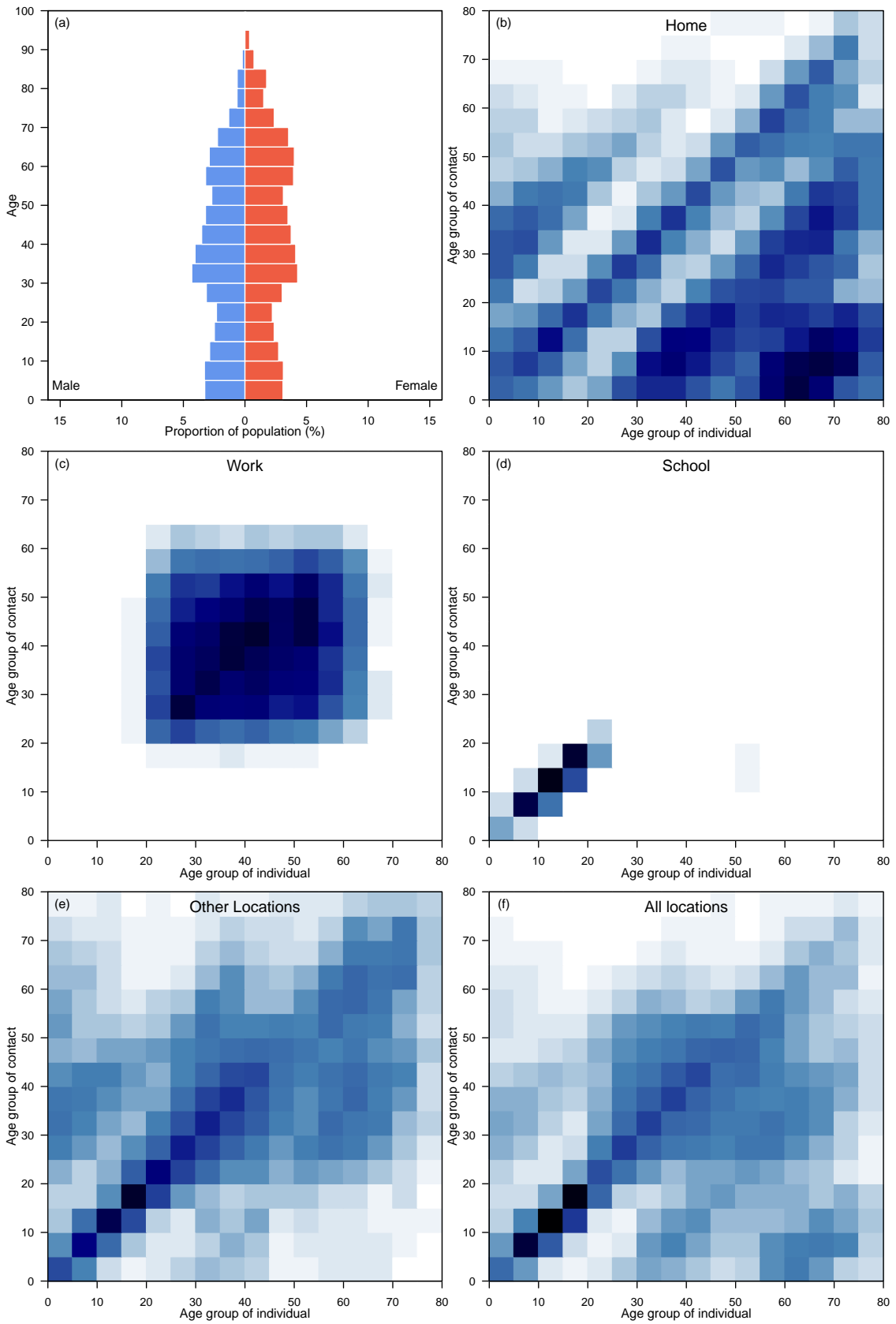
Rwanda
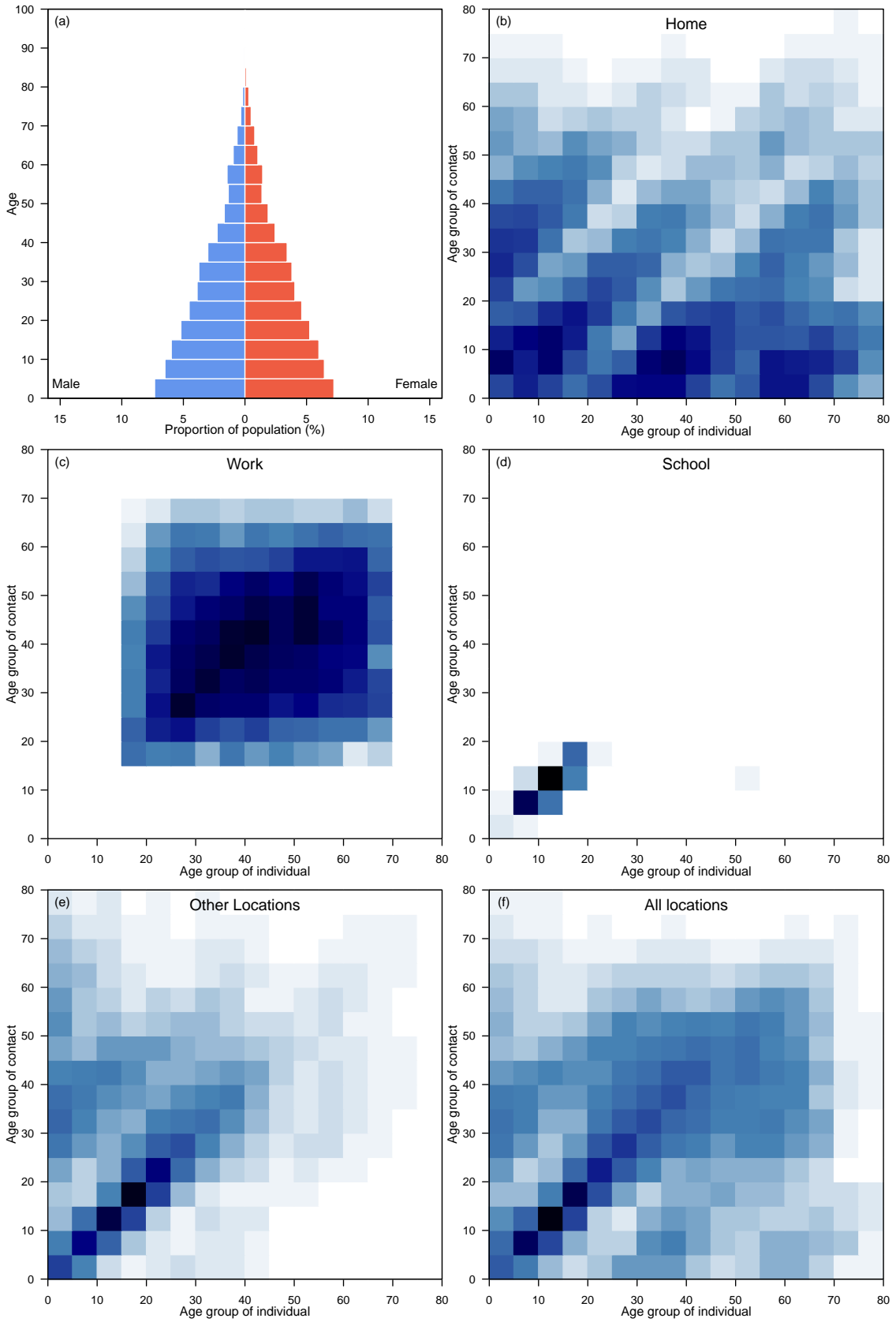
Saint Lucia
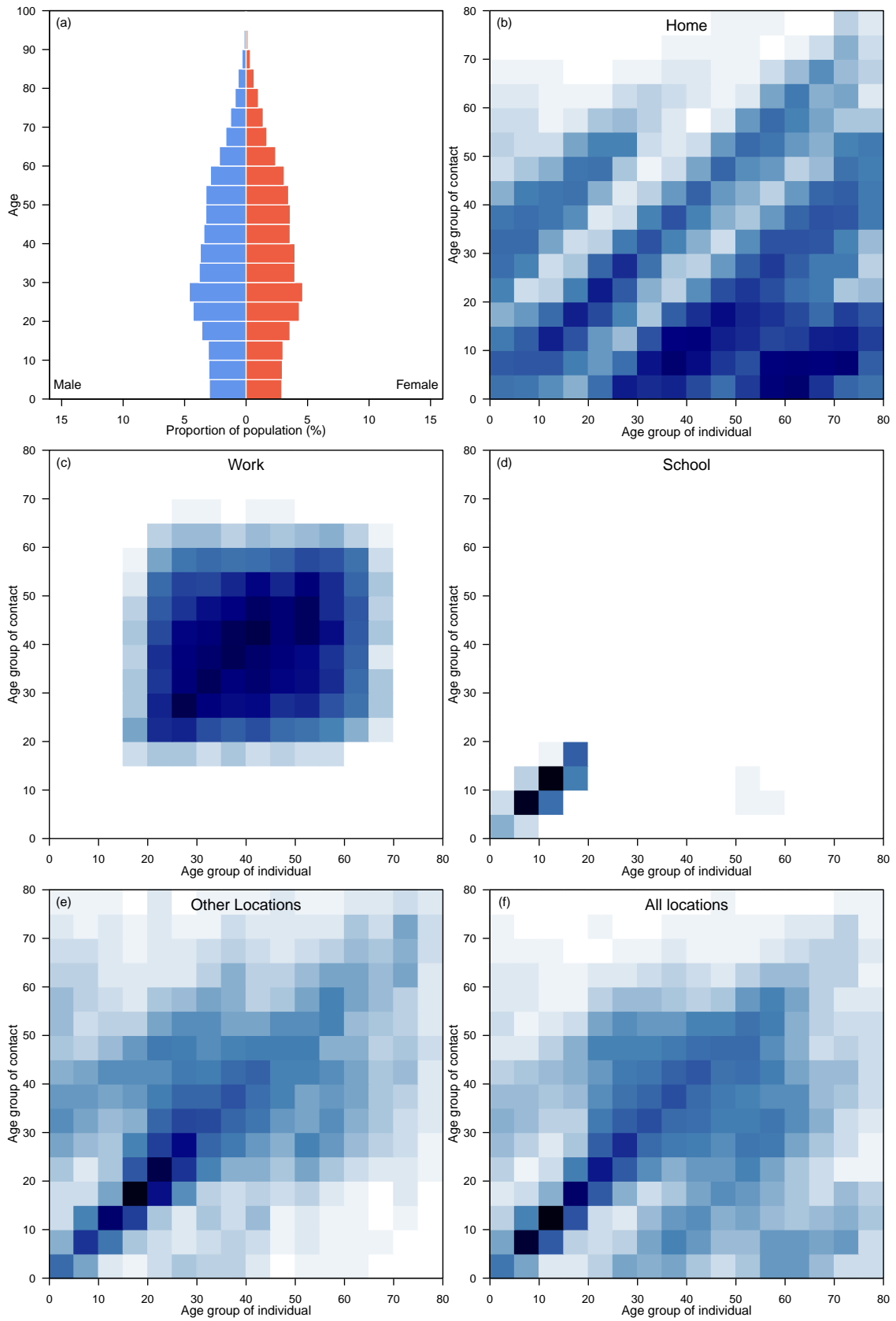
Saint Vincent and the Grenadines
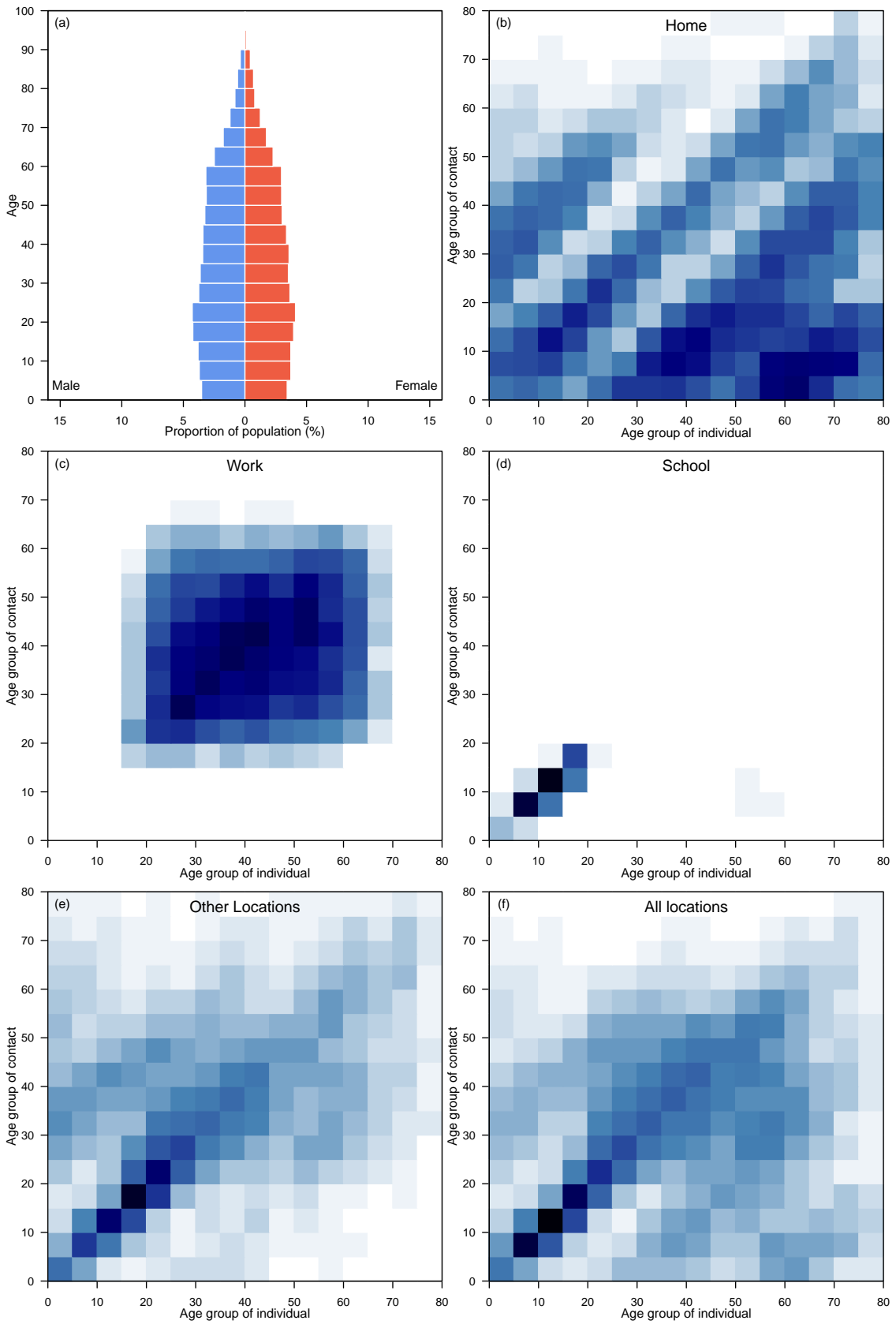
Samoa
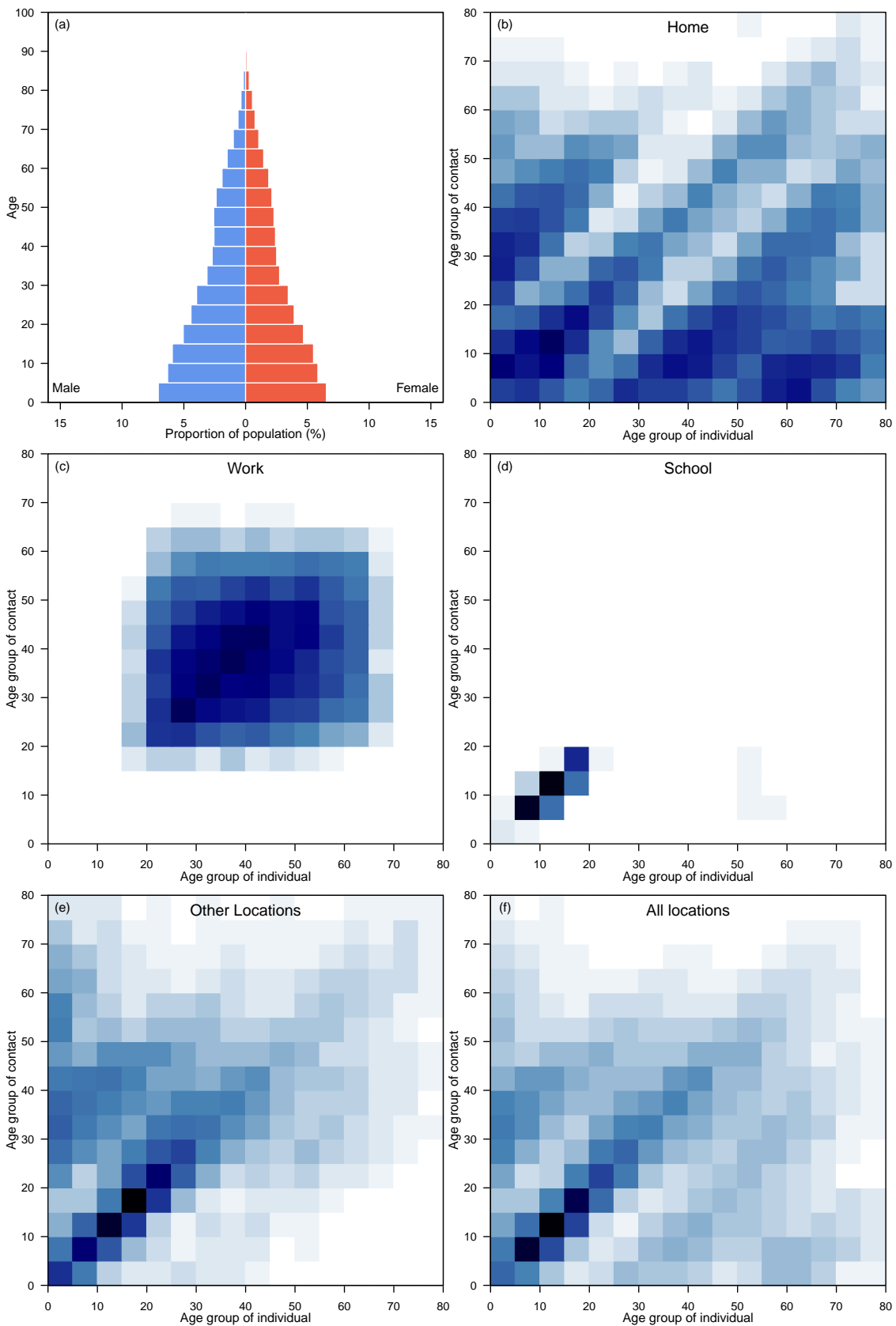
Sao Tome and Principe
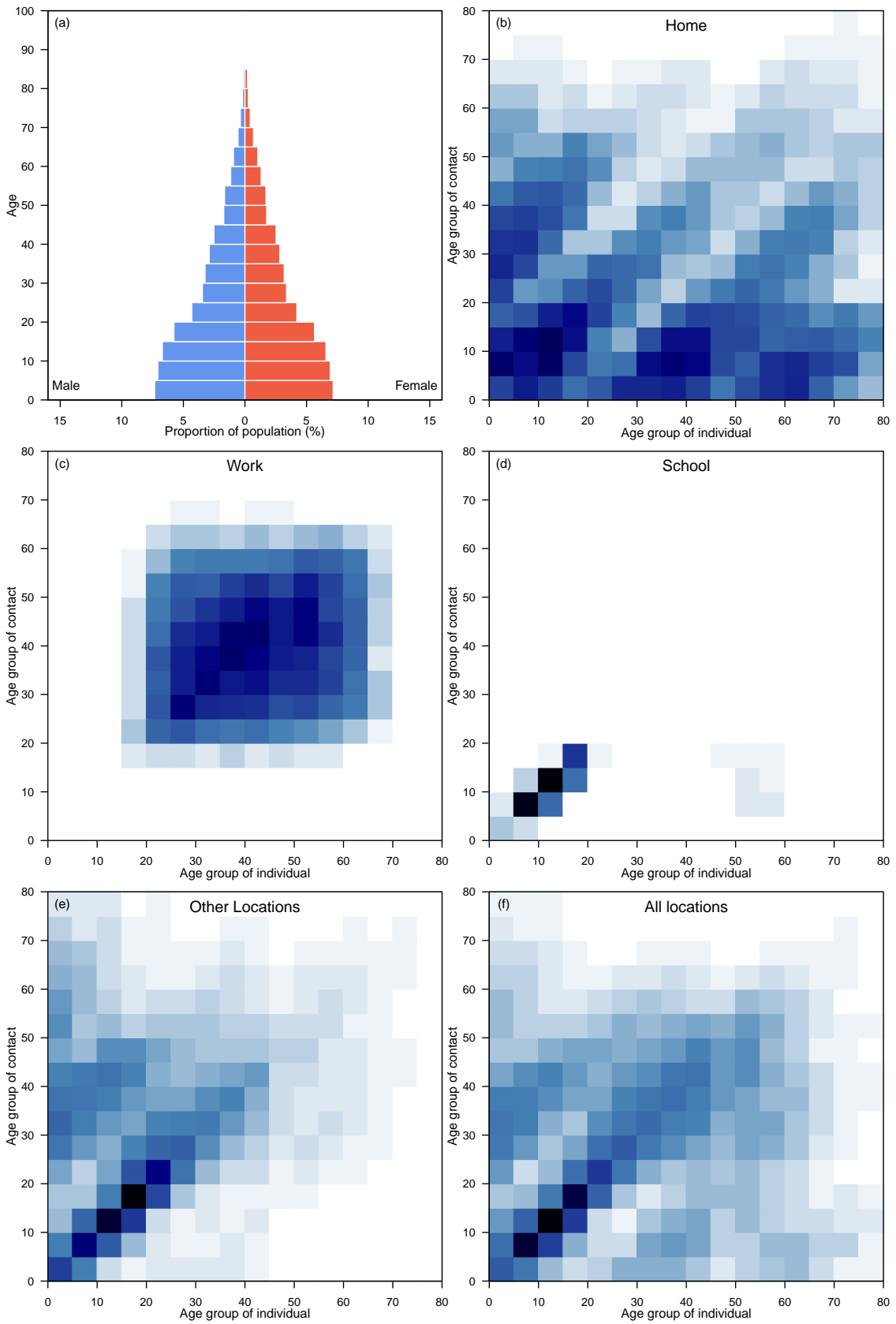
Saudi Arabia
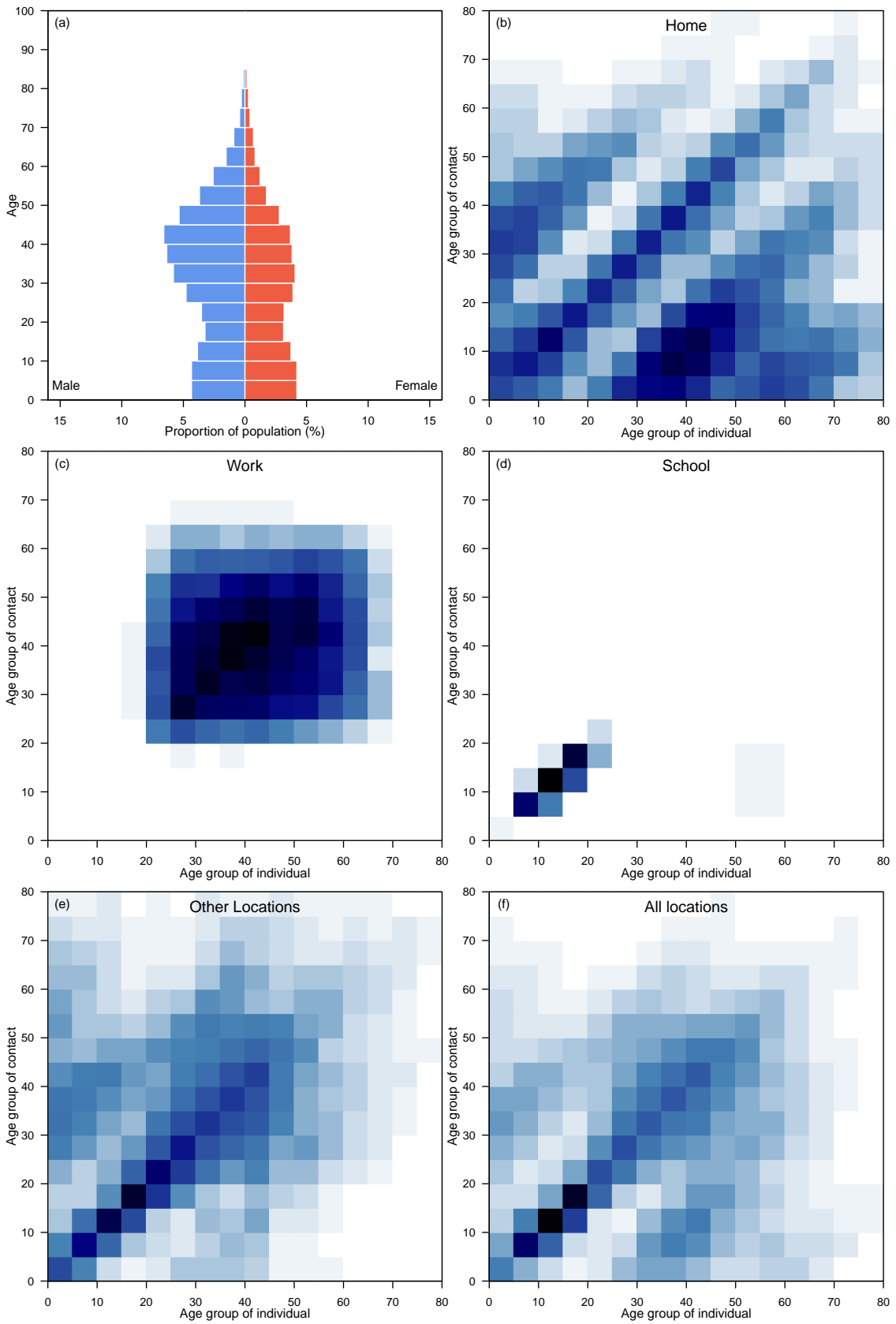
Senegal
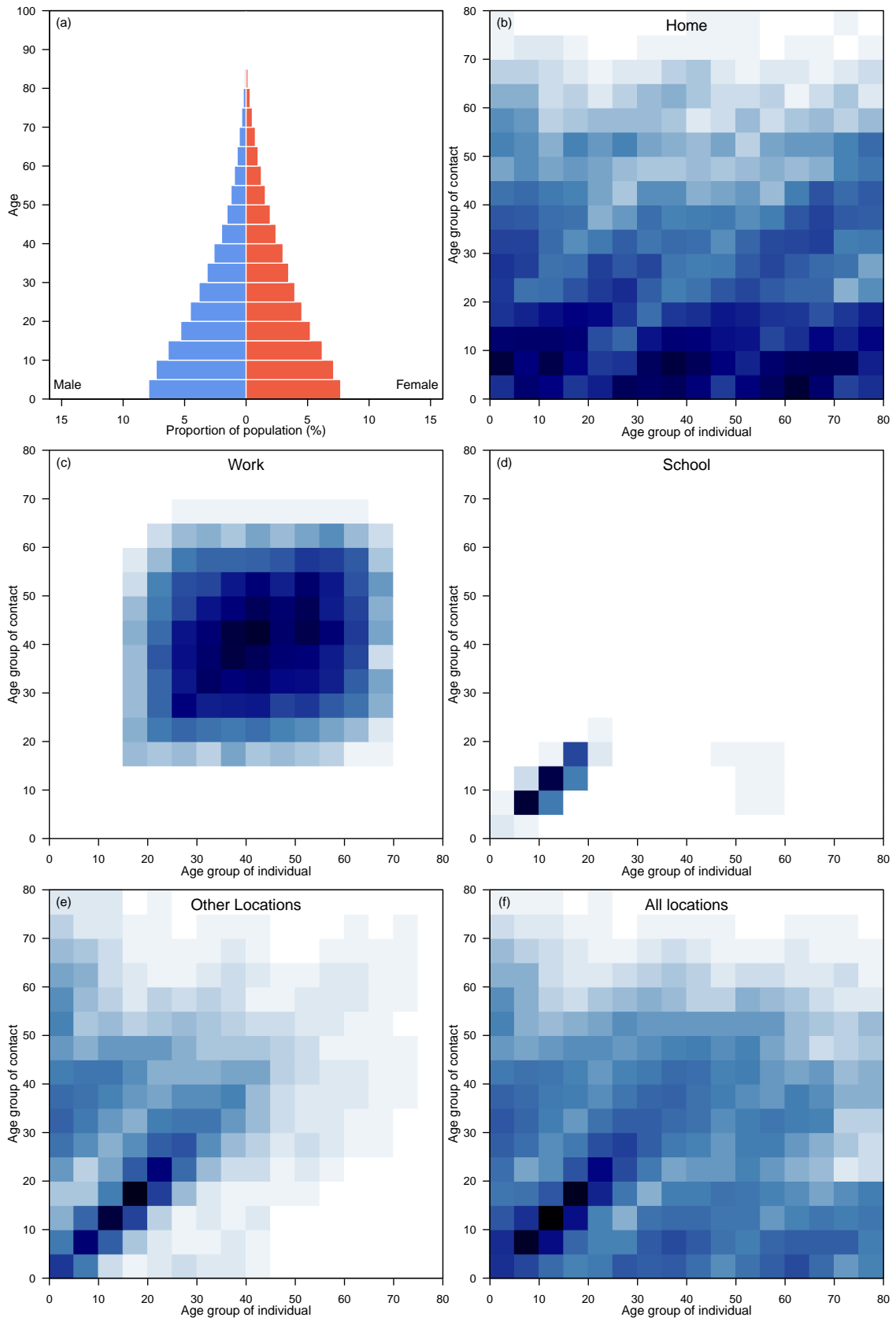
Serbia
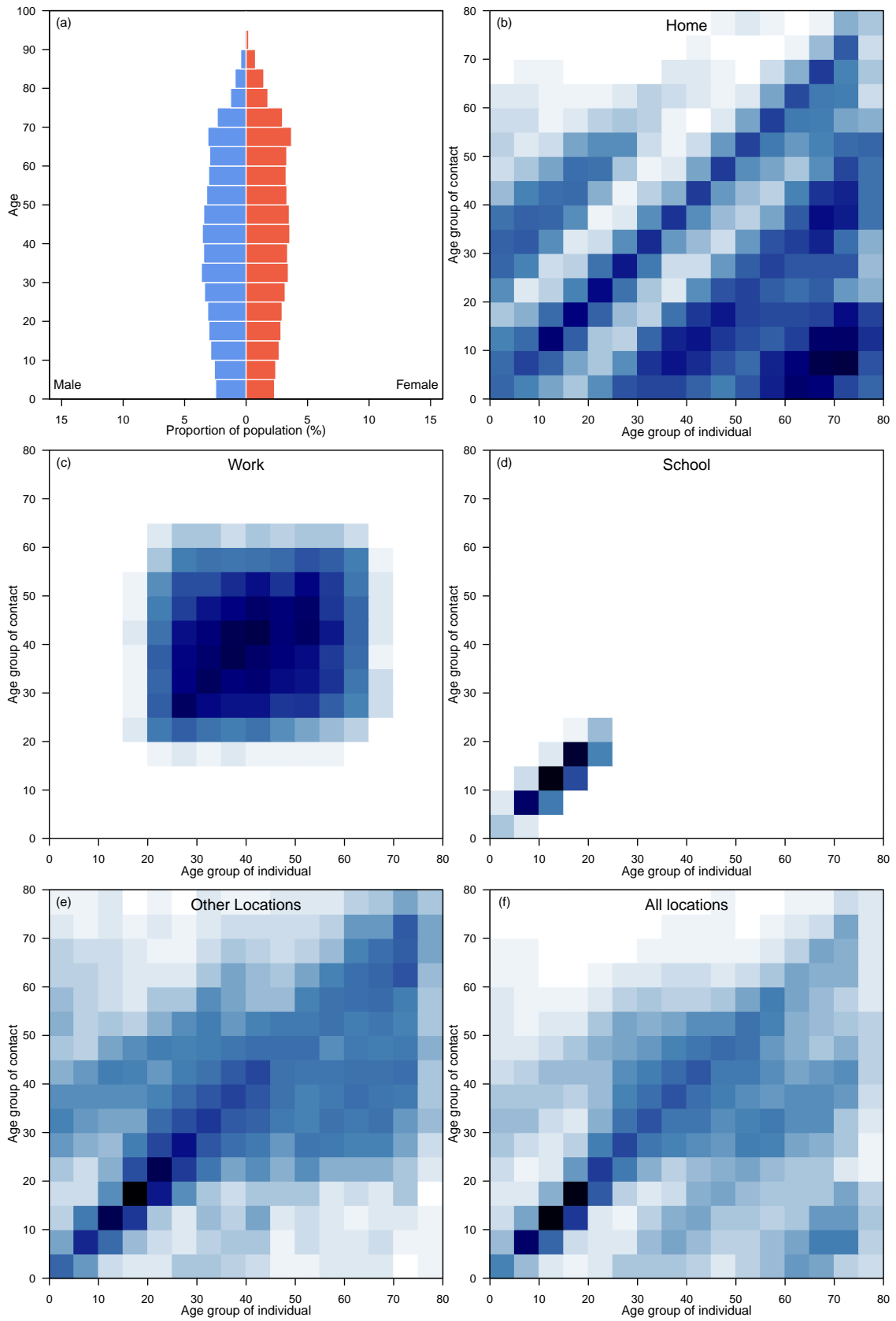
Sierra Leone
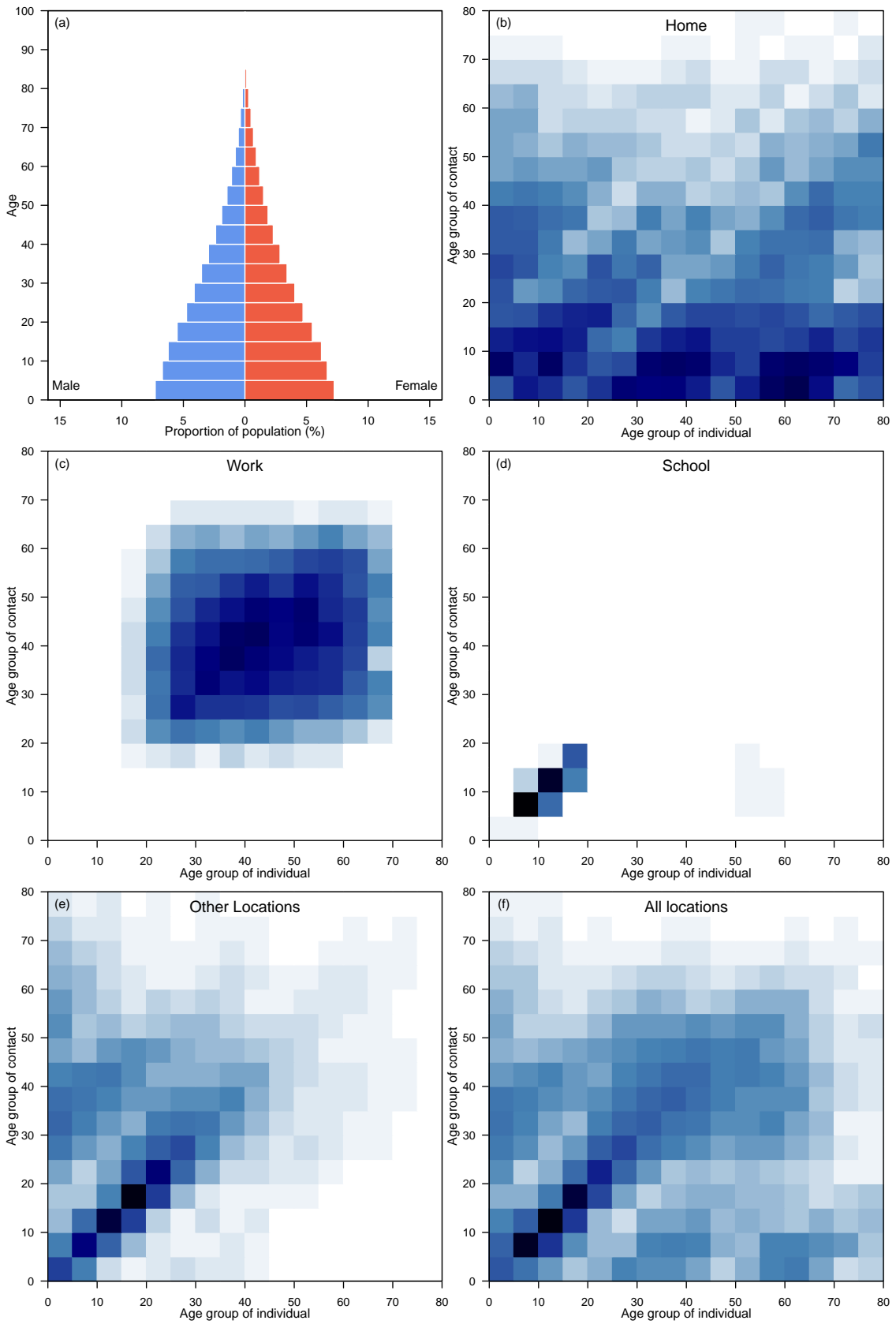
Singapore
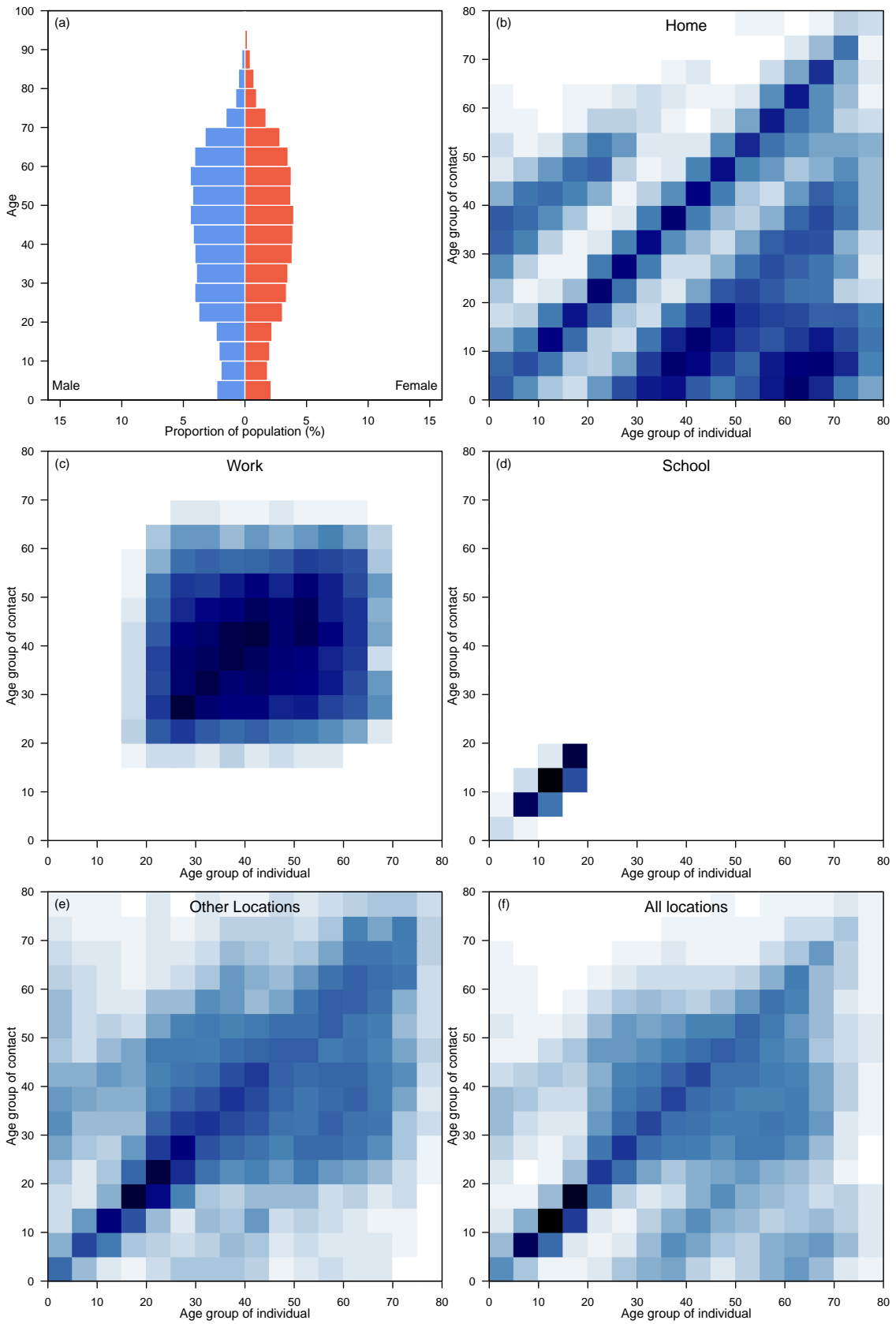
Slovakia
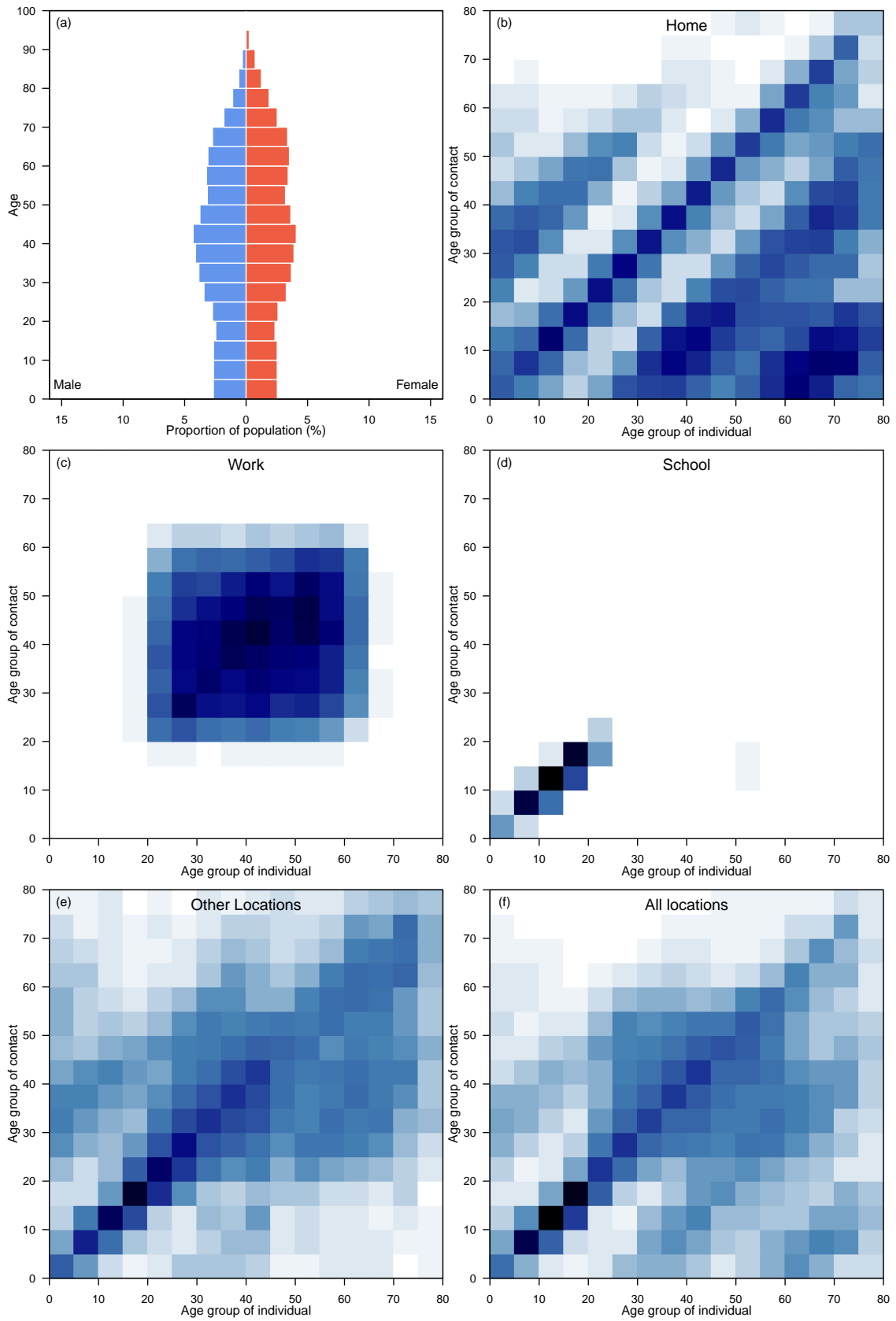
Slovenia
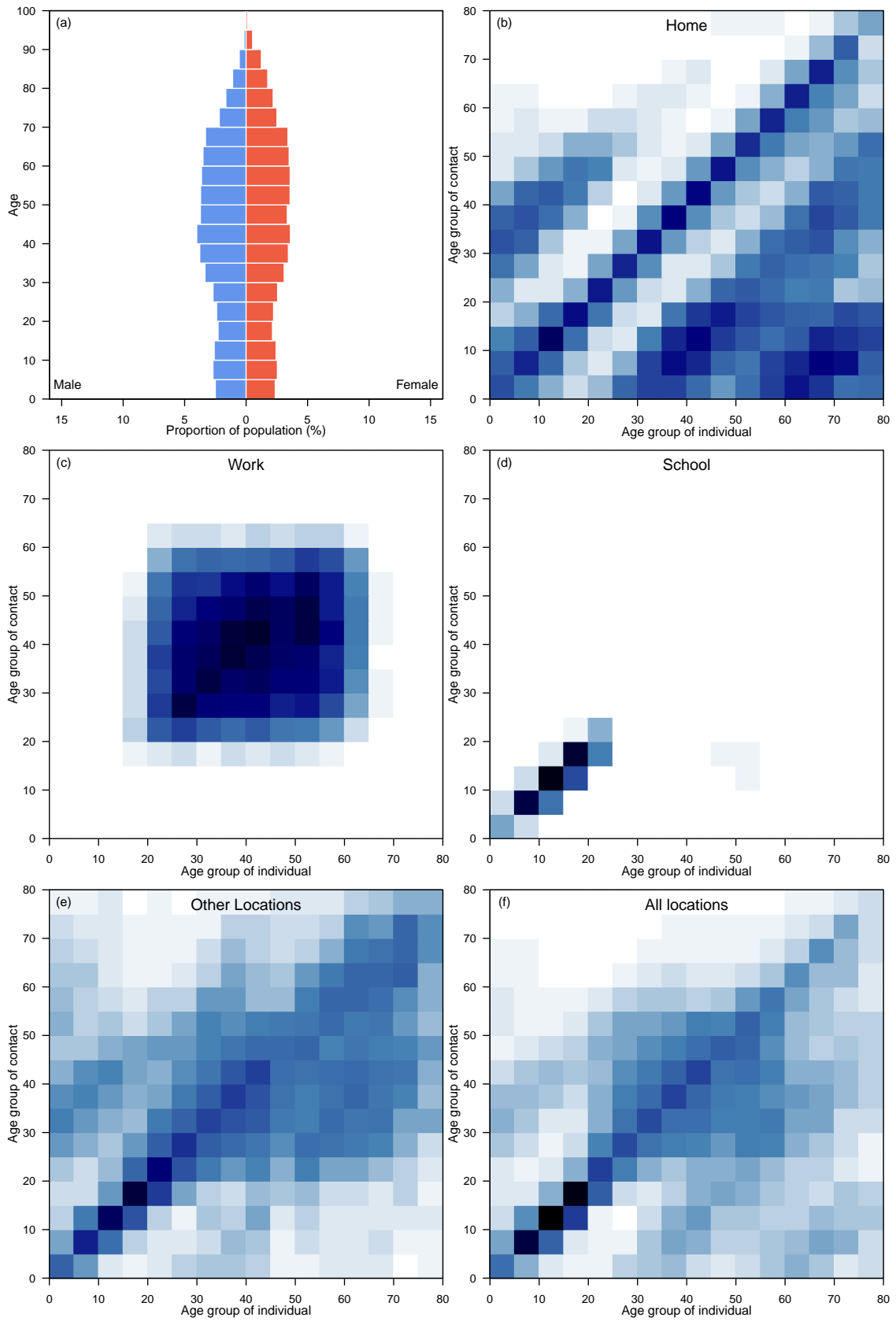
Solomon Islands
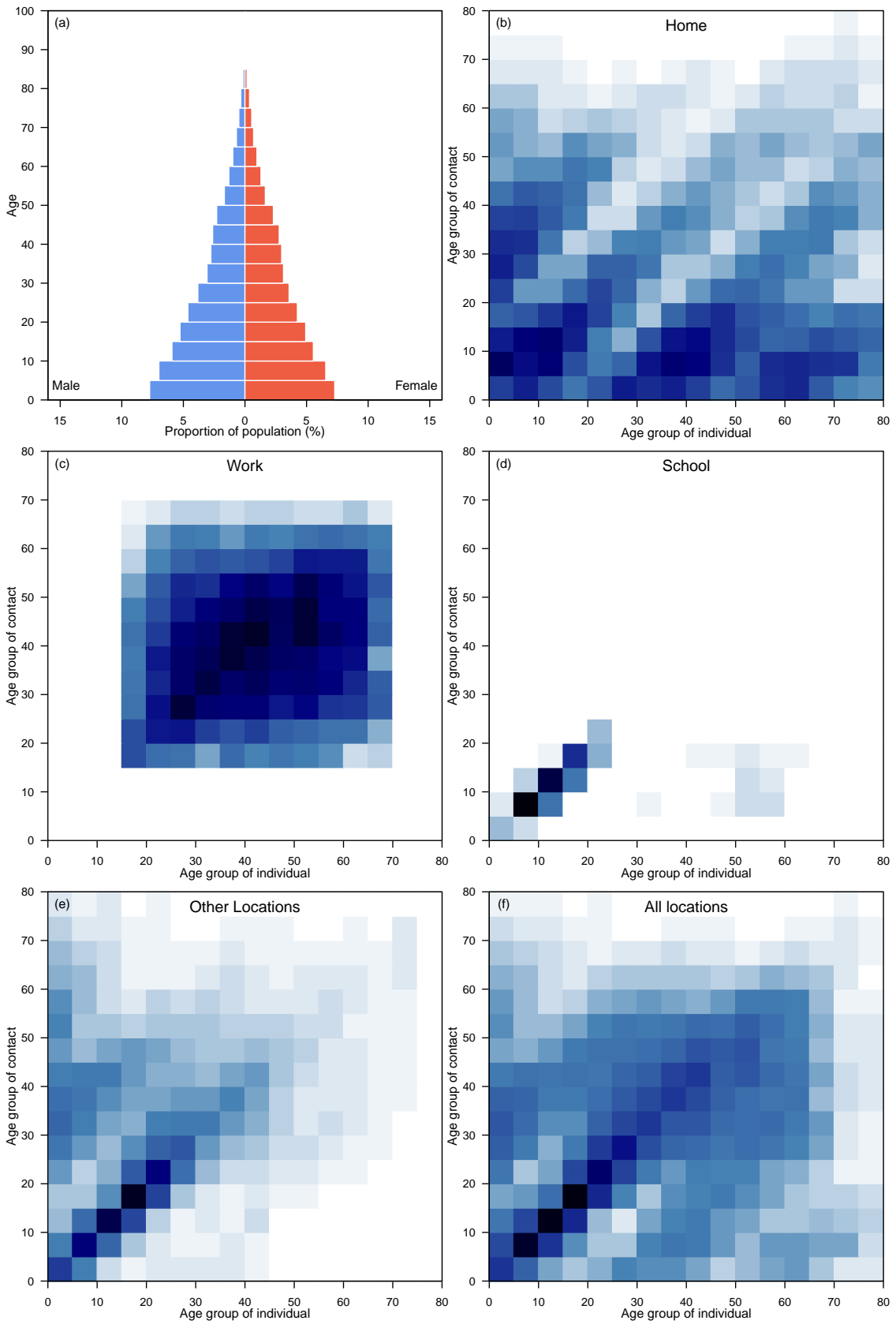
South Africa
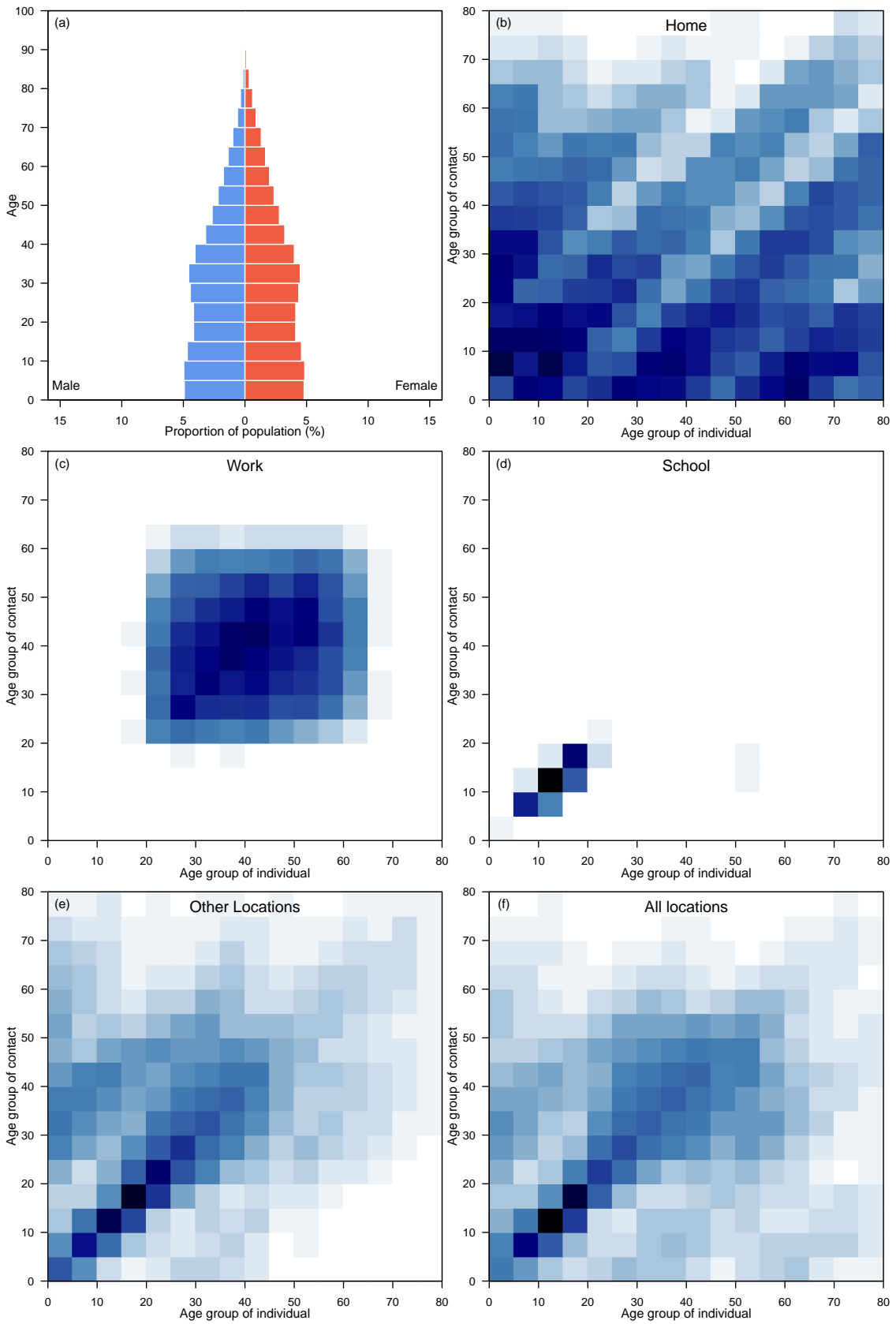
South Sudan
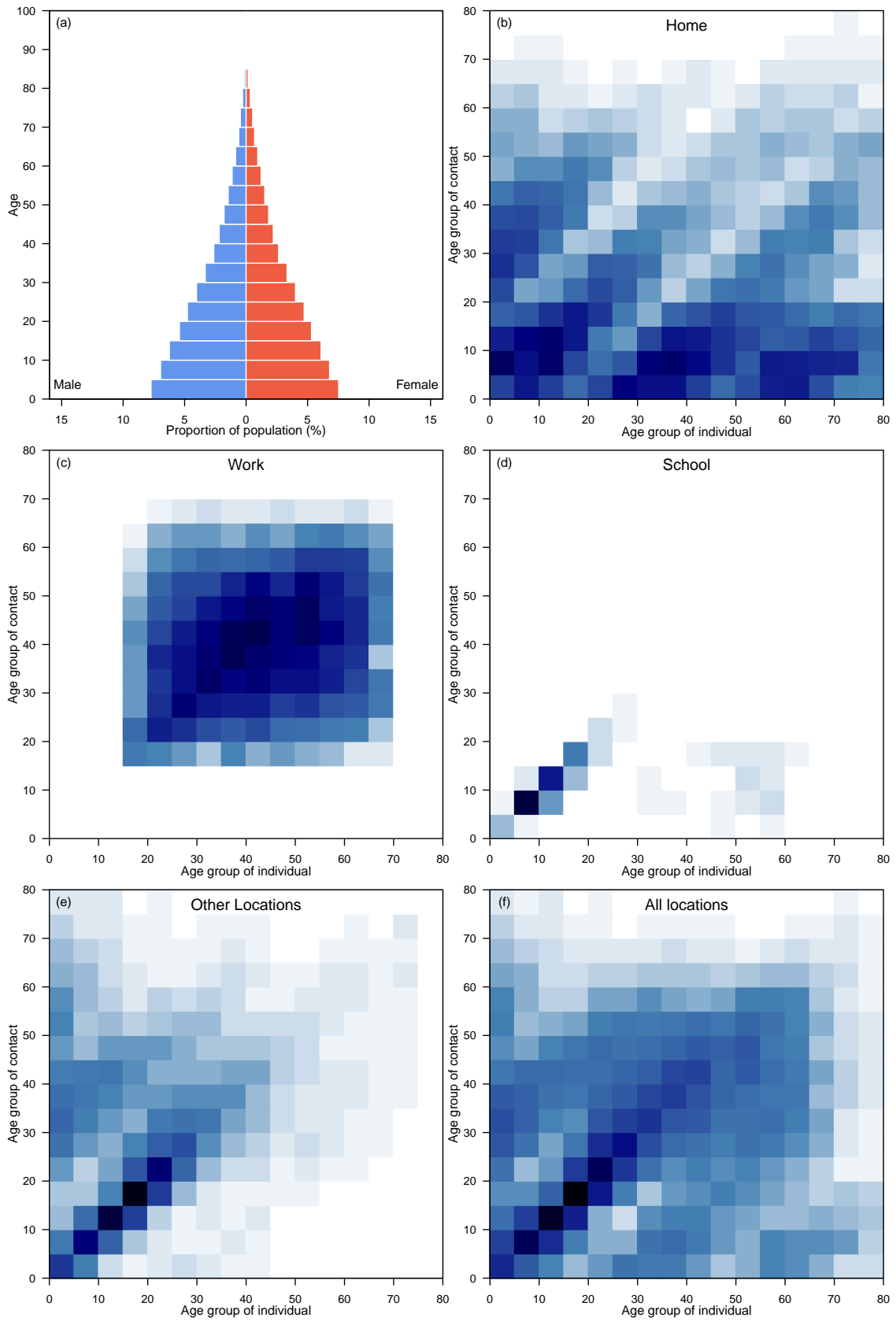
Spain
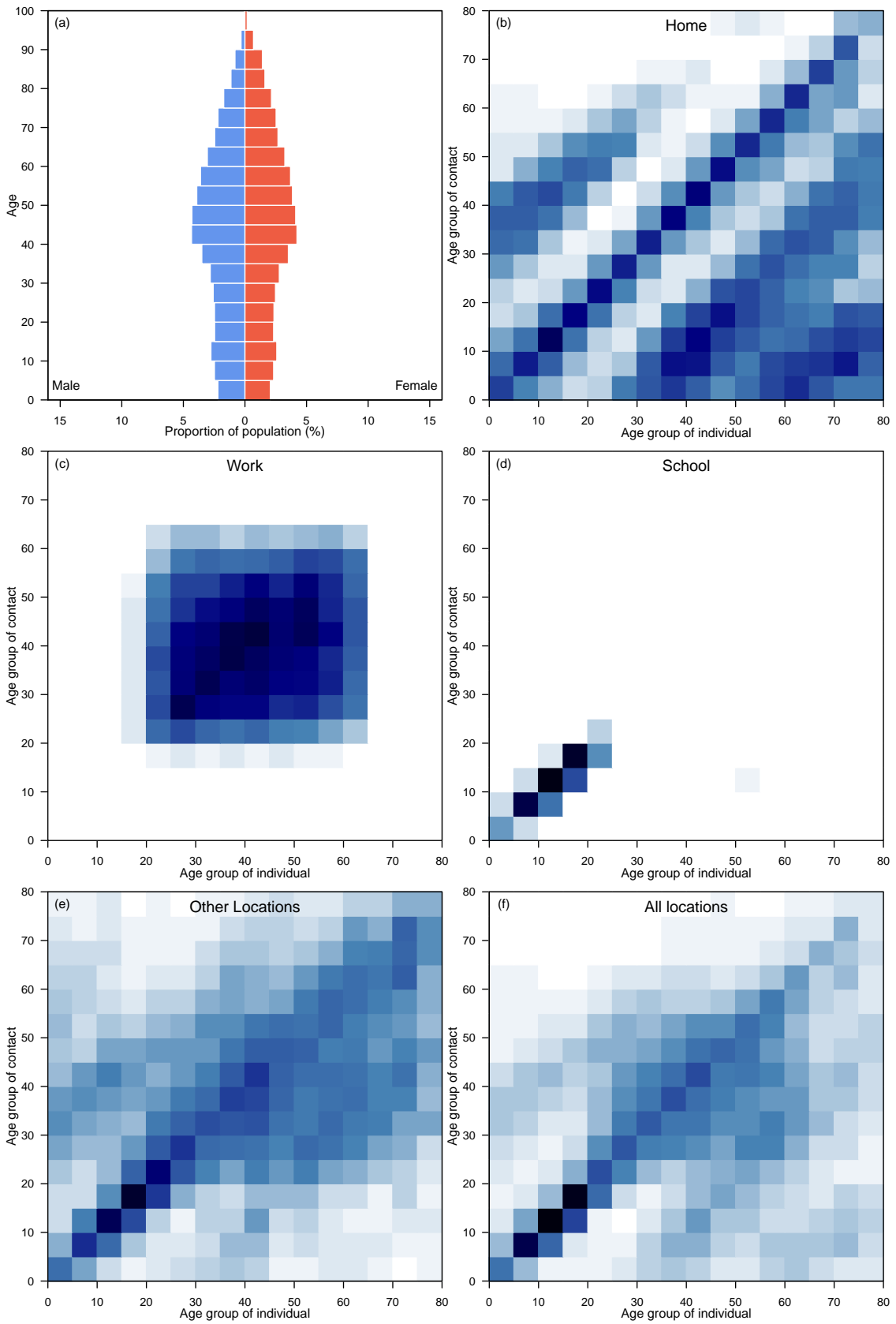
Sri Lanka
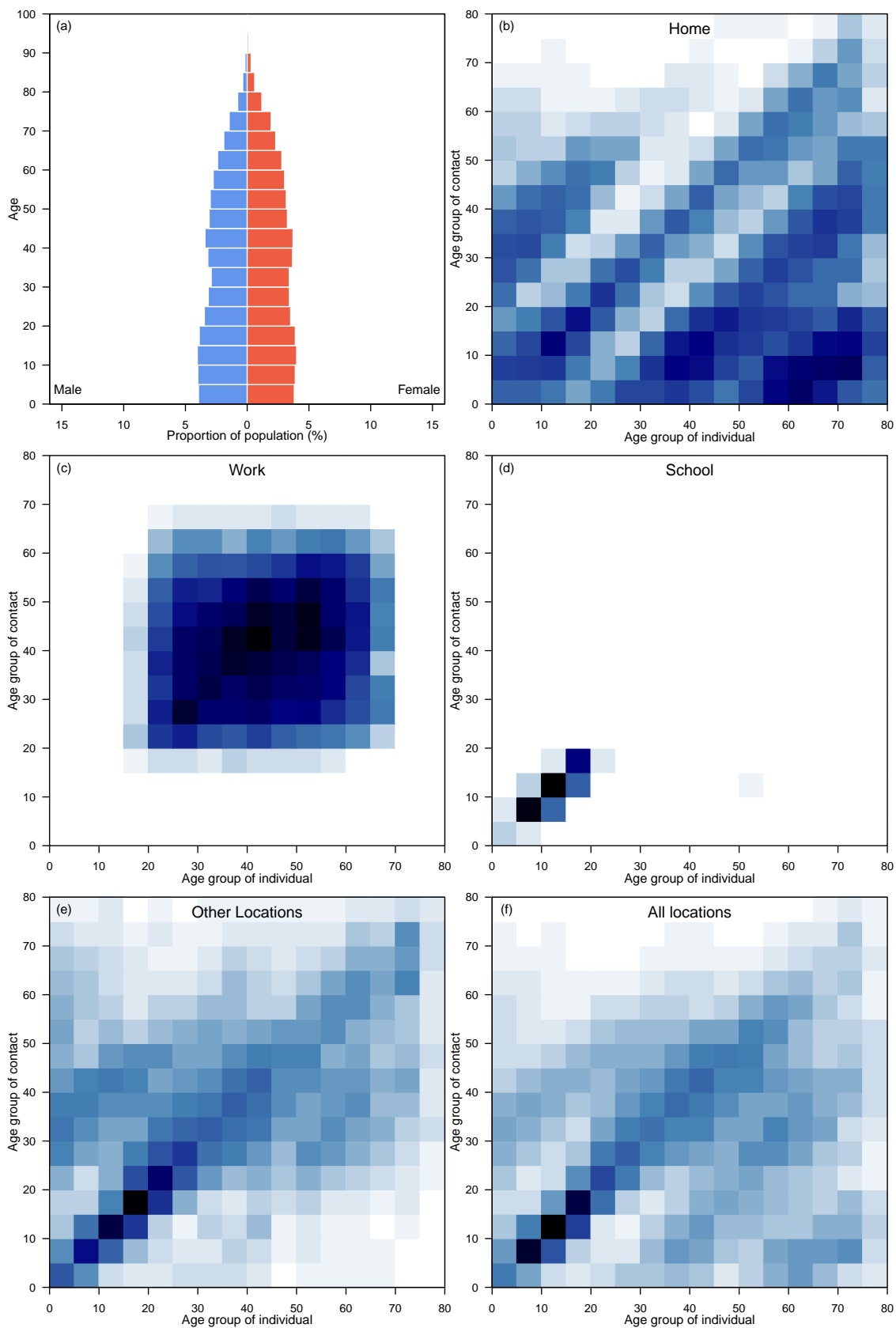
State of Palestine
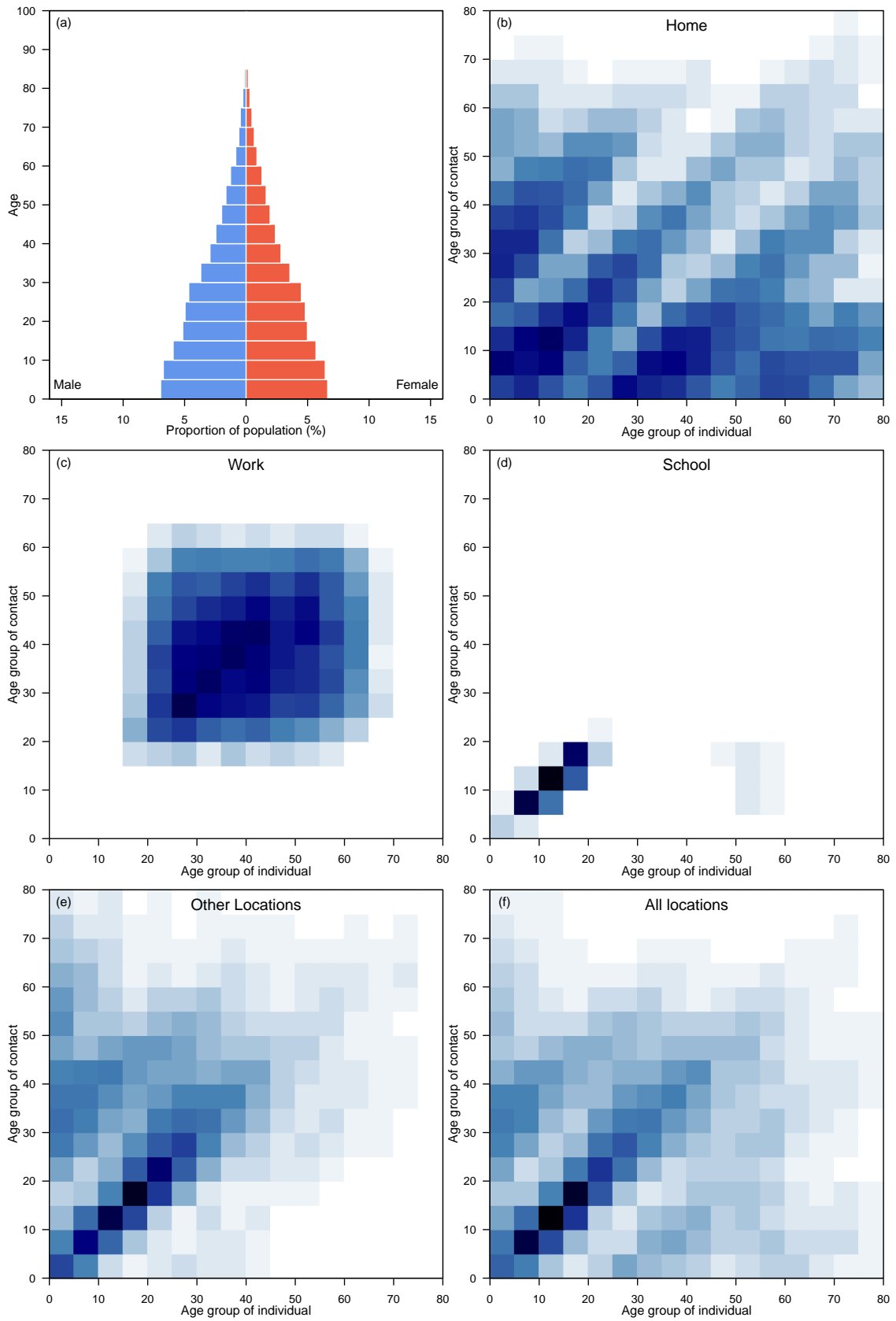
Sudan
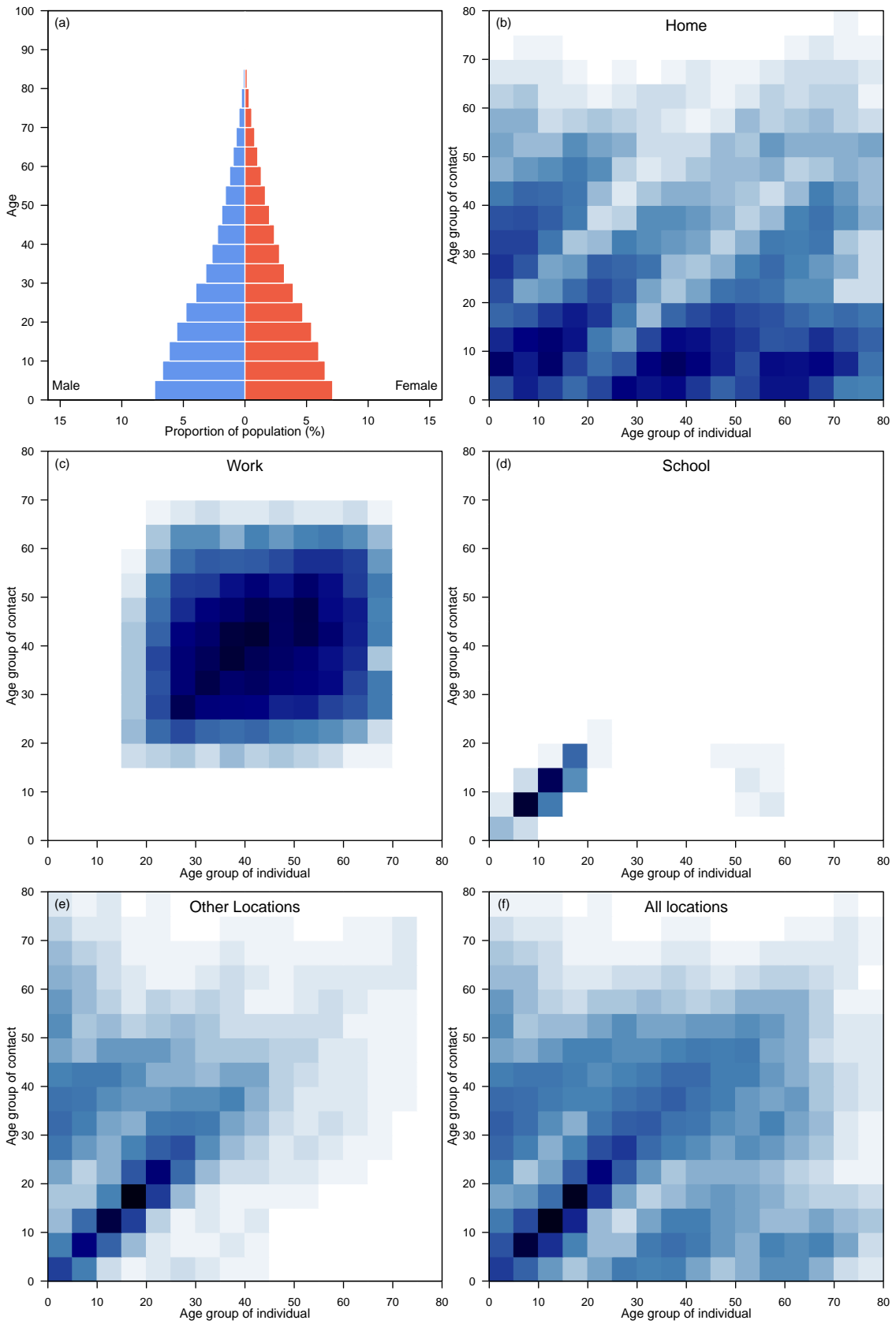
Suriname
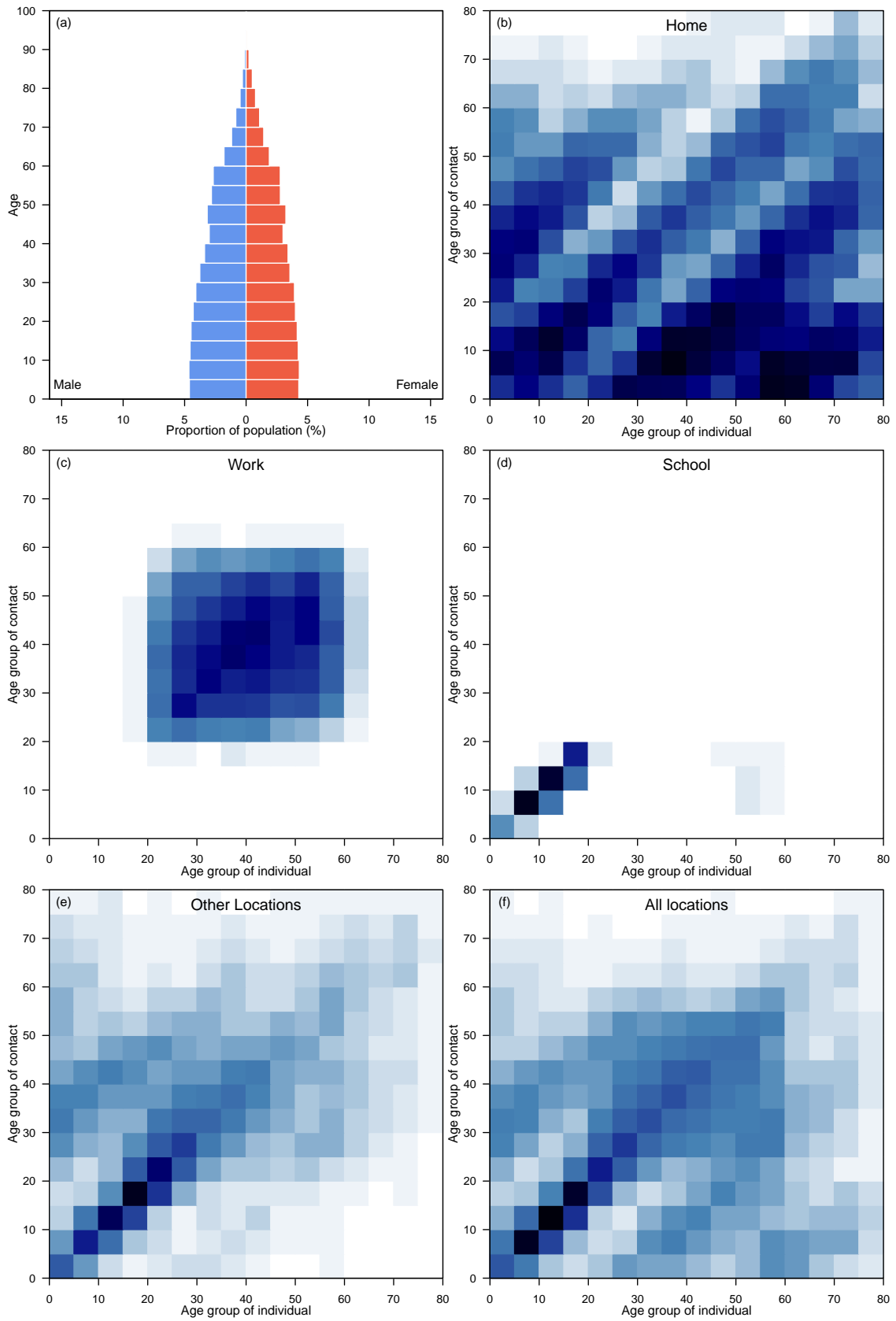
Sweden
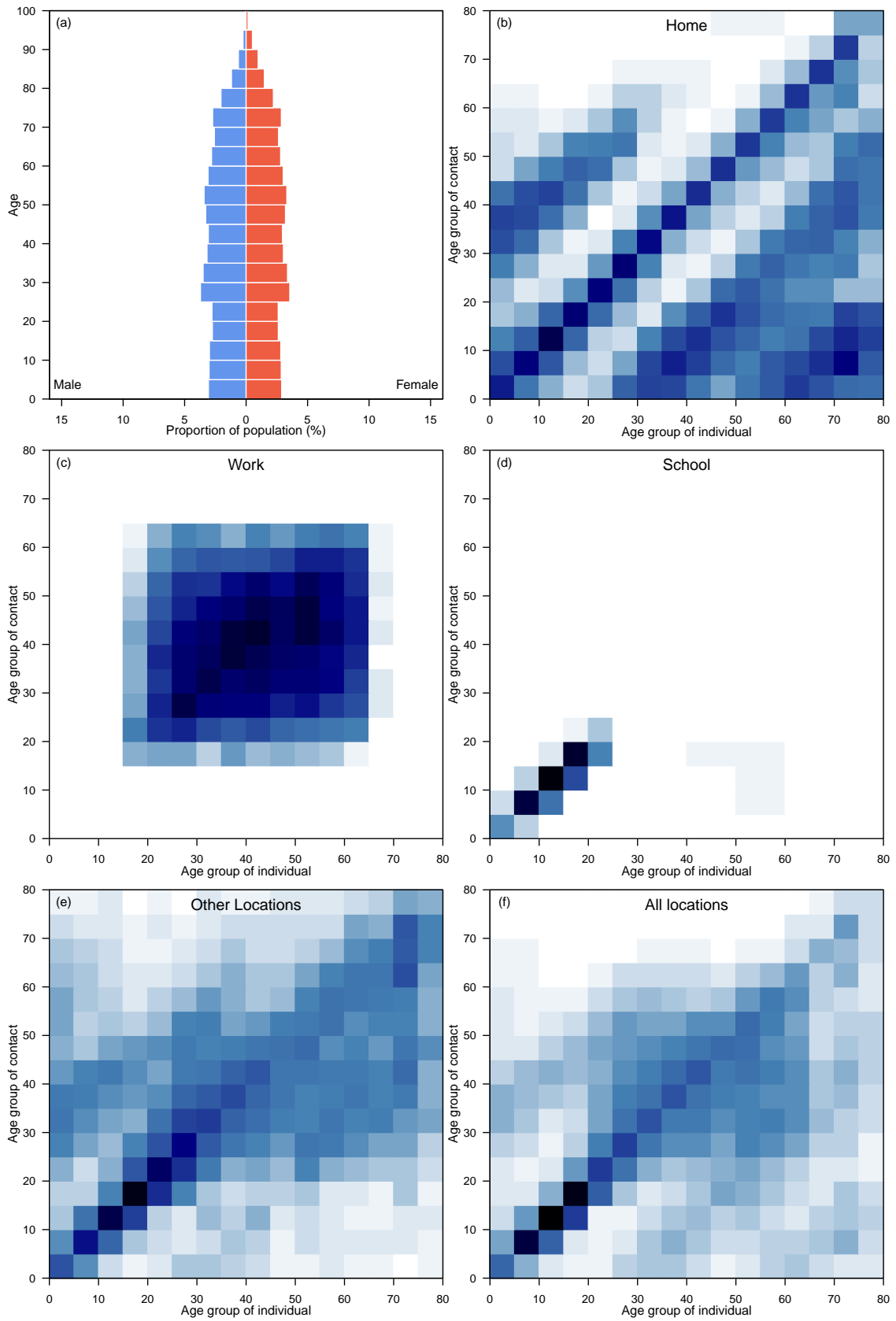
Switzerland
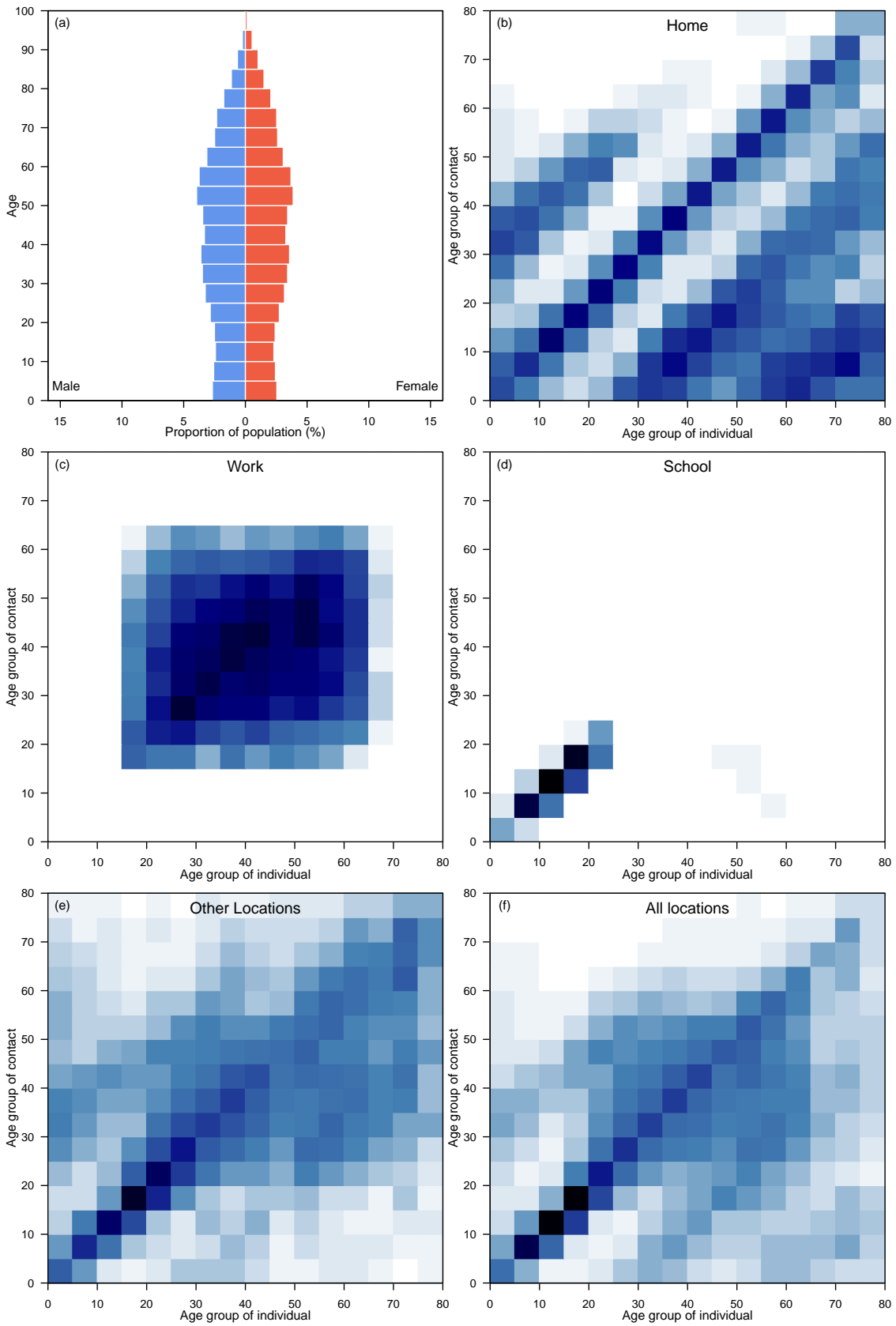
Syrian Arab Republic
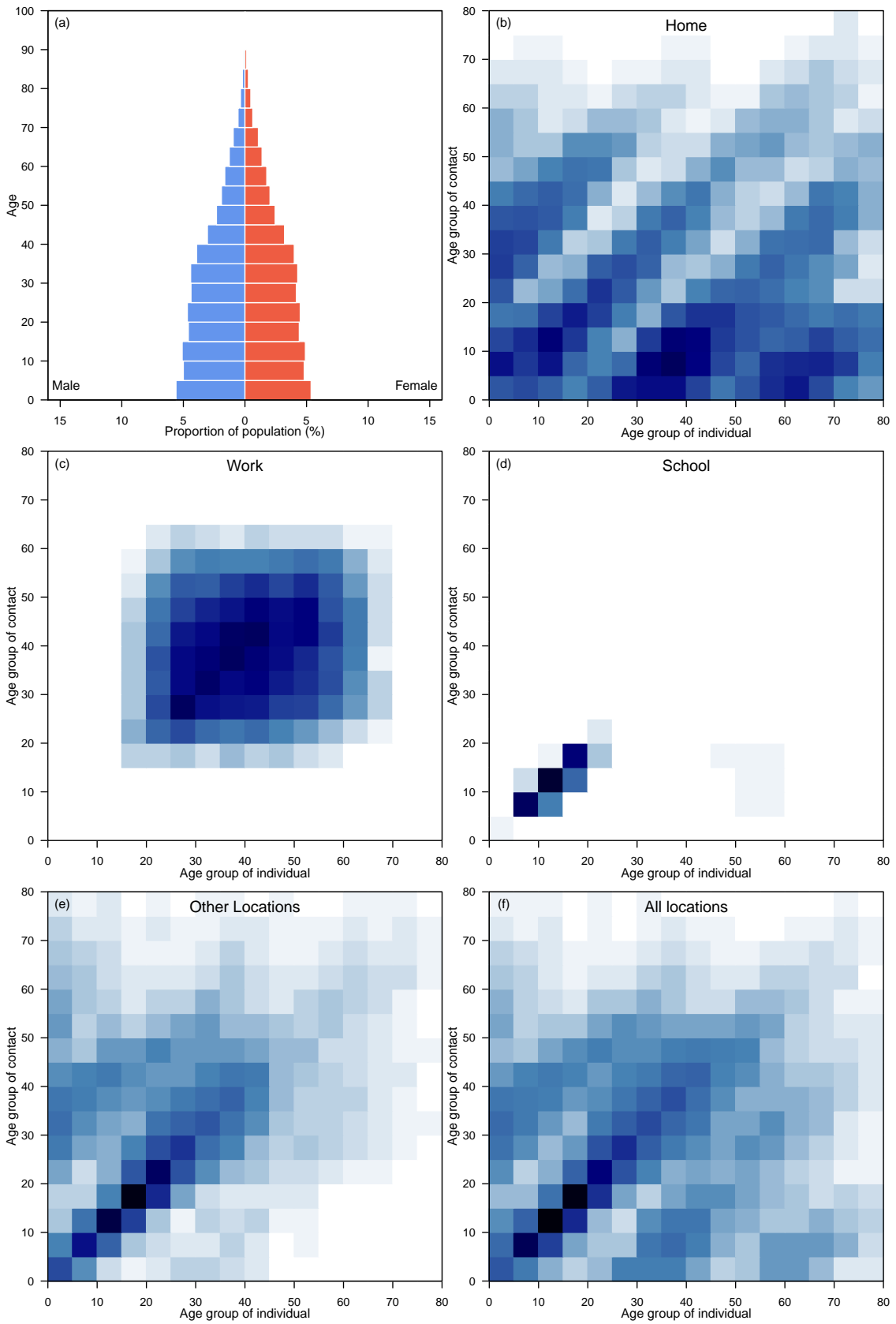
Tajikistan
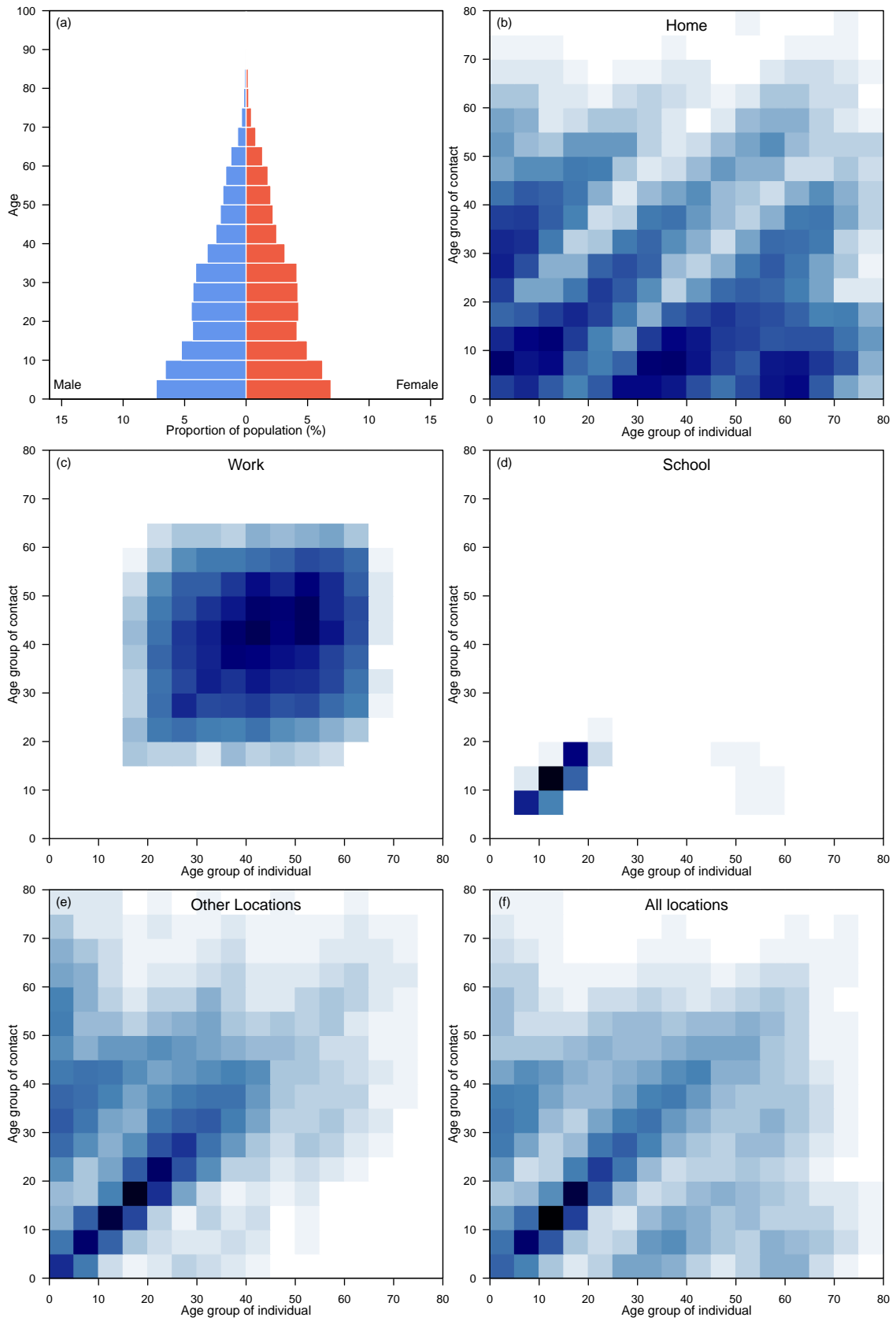
Thailand
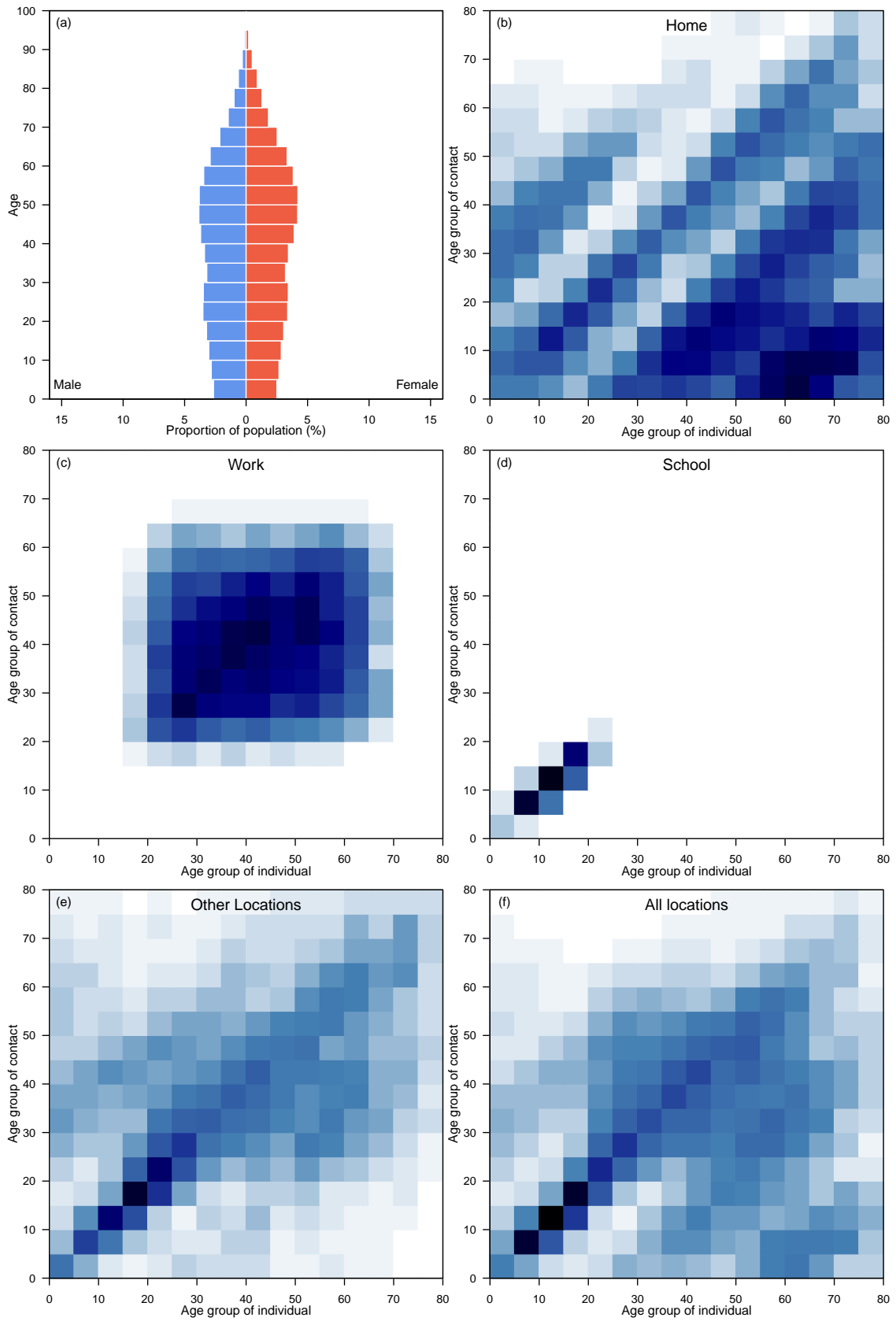
Timor−Leste
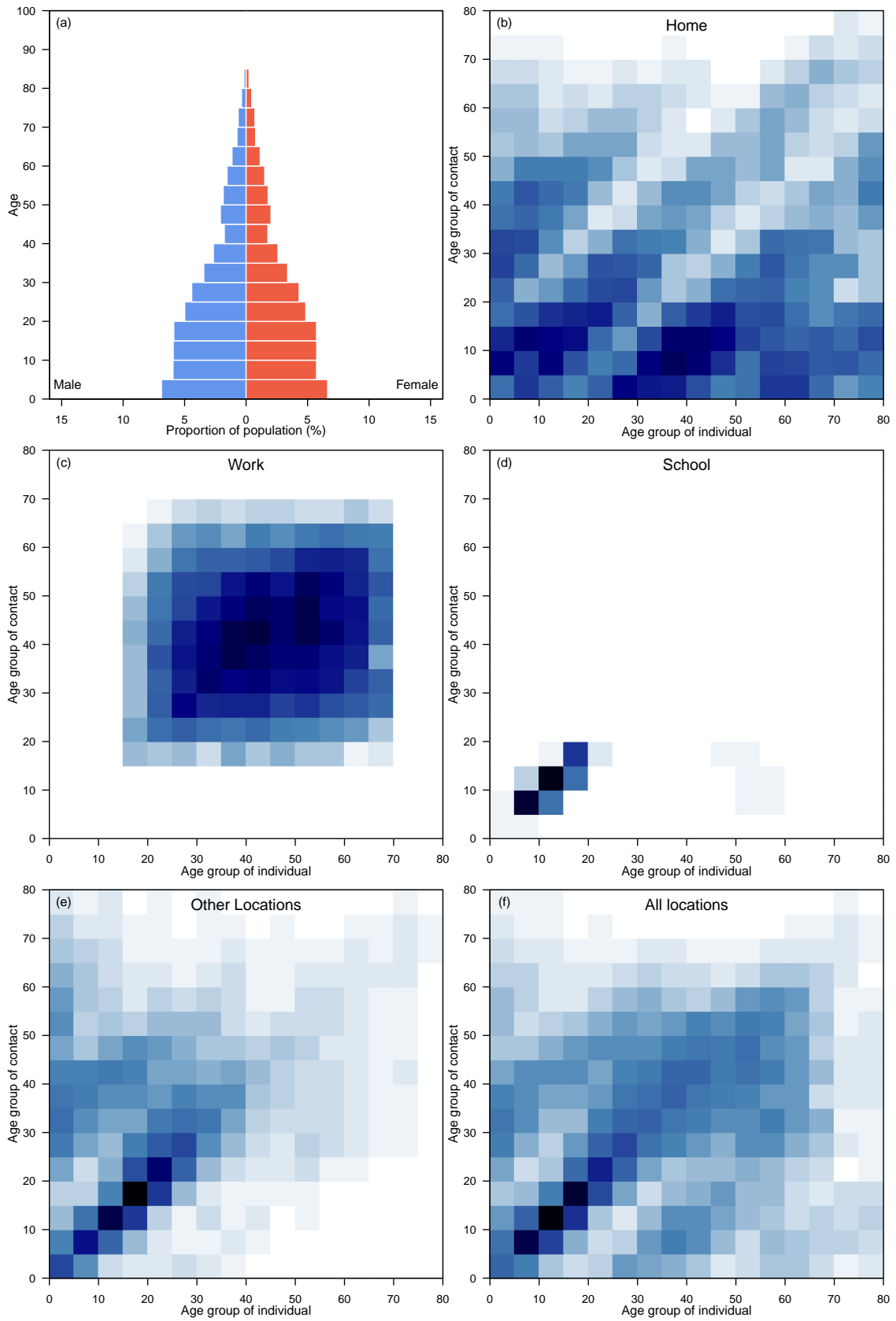
Togo
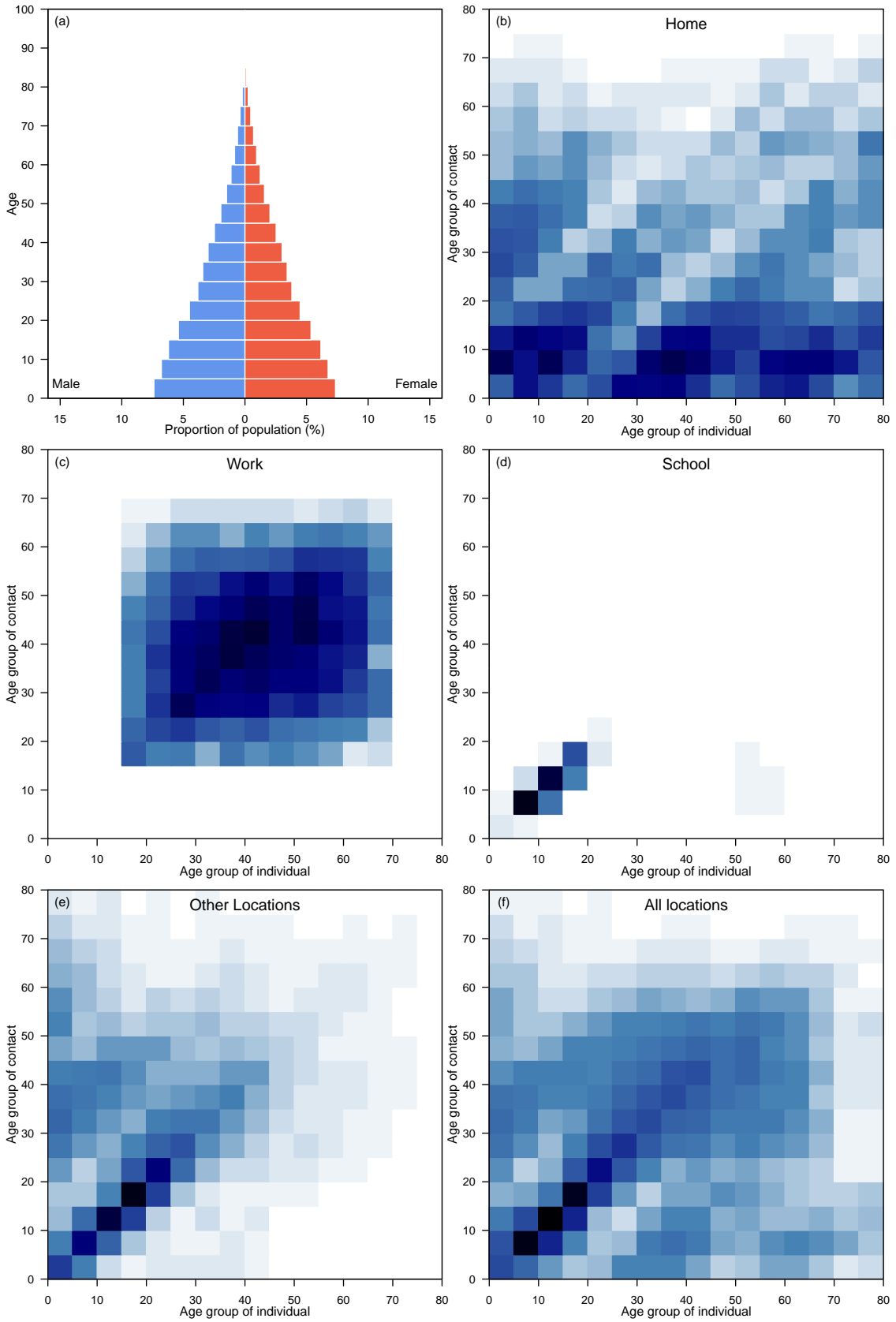
Tonga
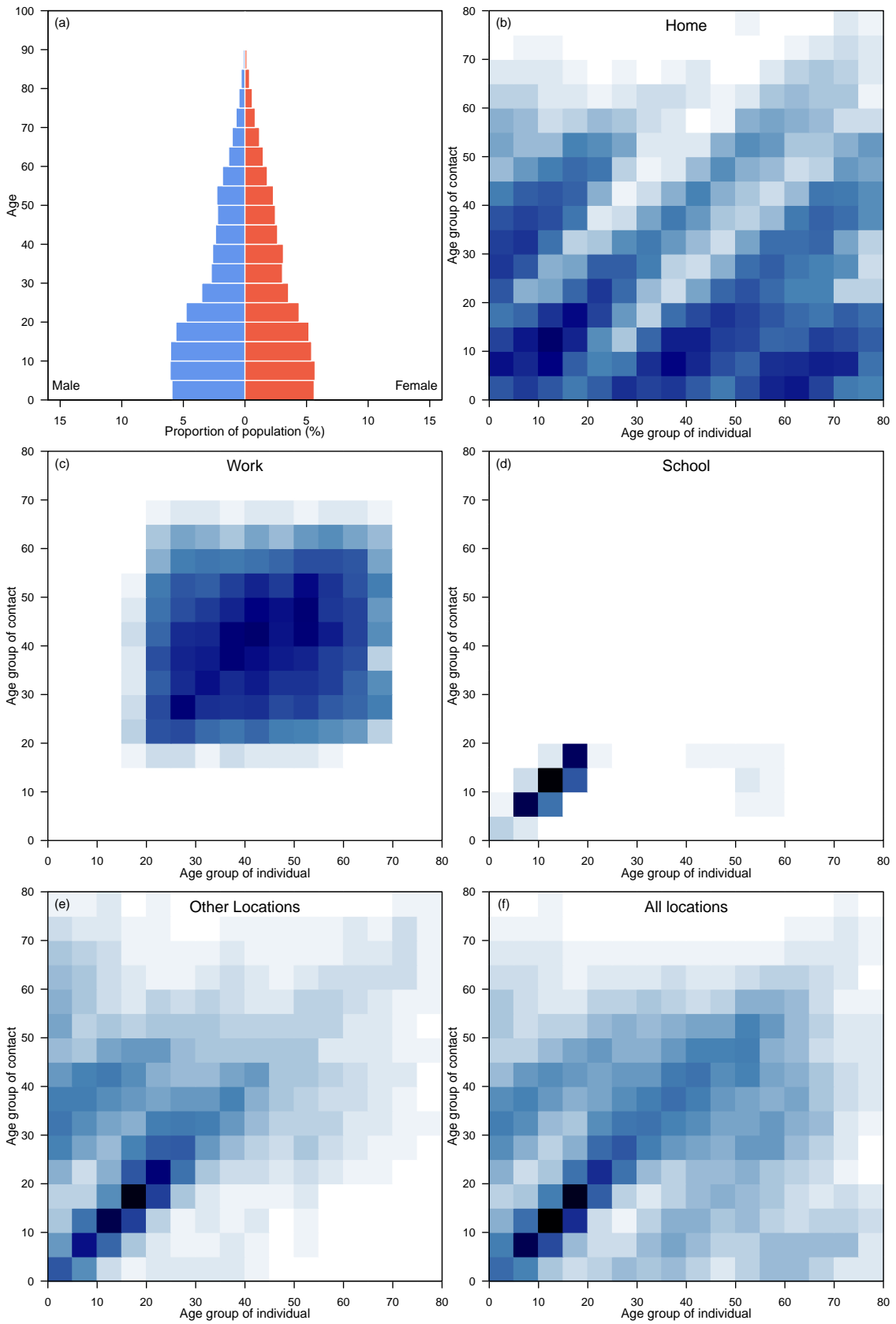
Trinidad and Tobago
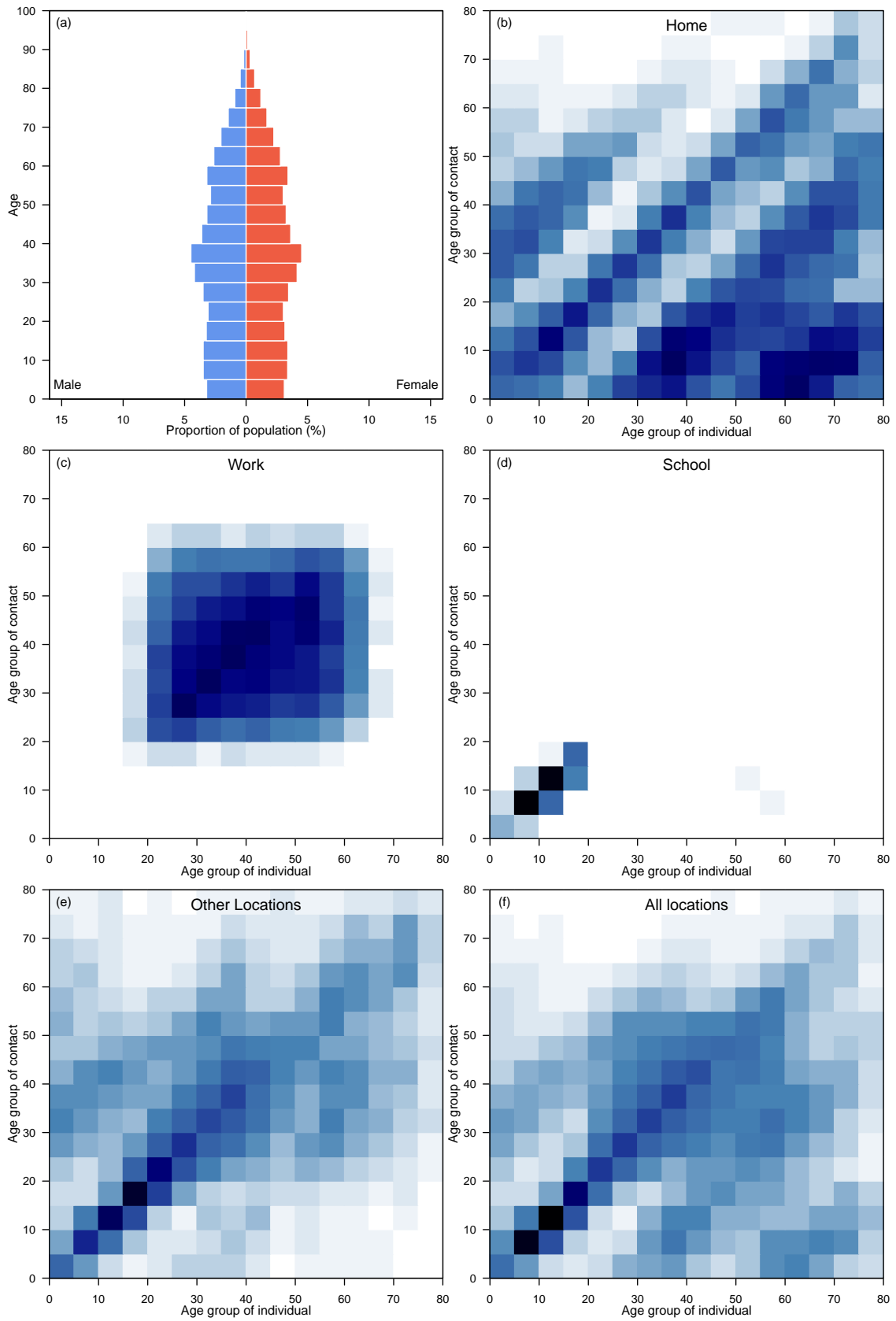
Tunisia
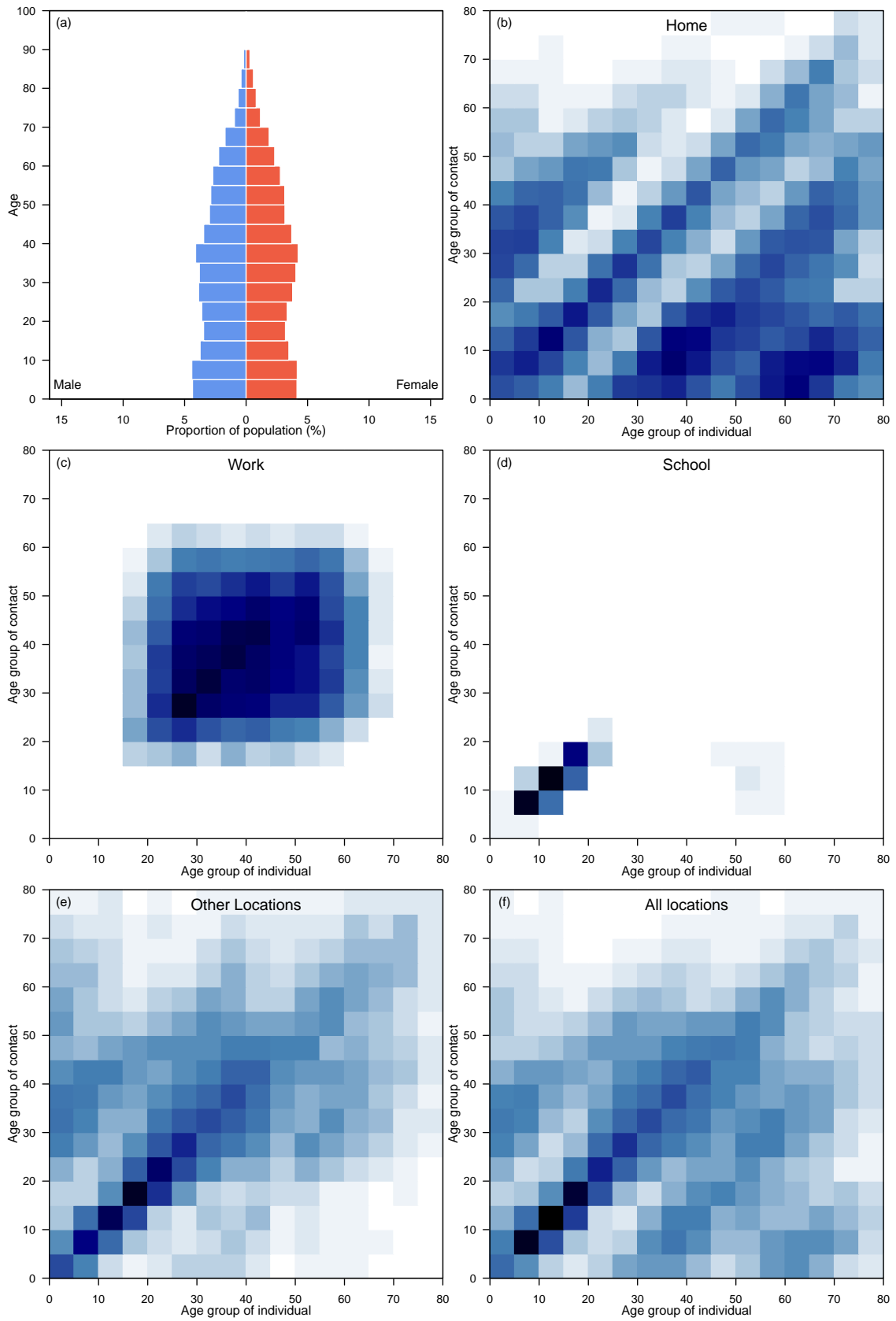
Turkey
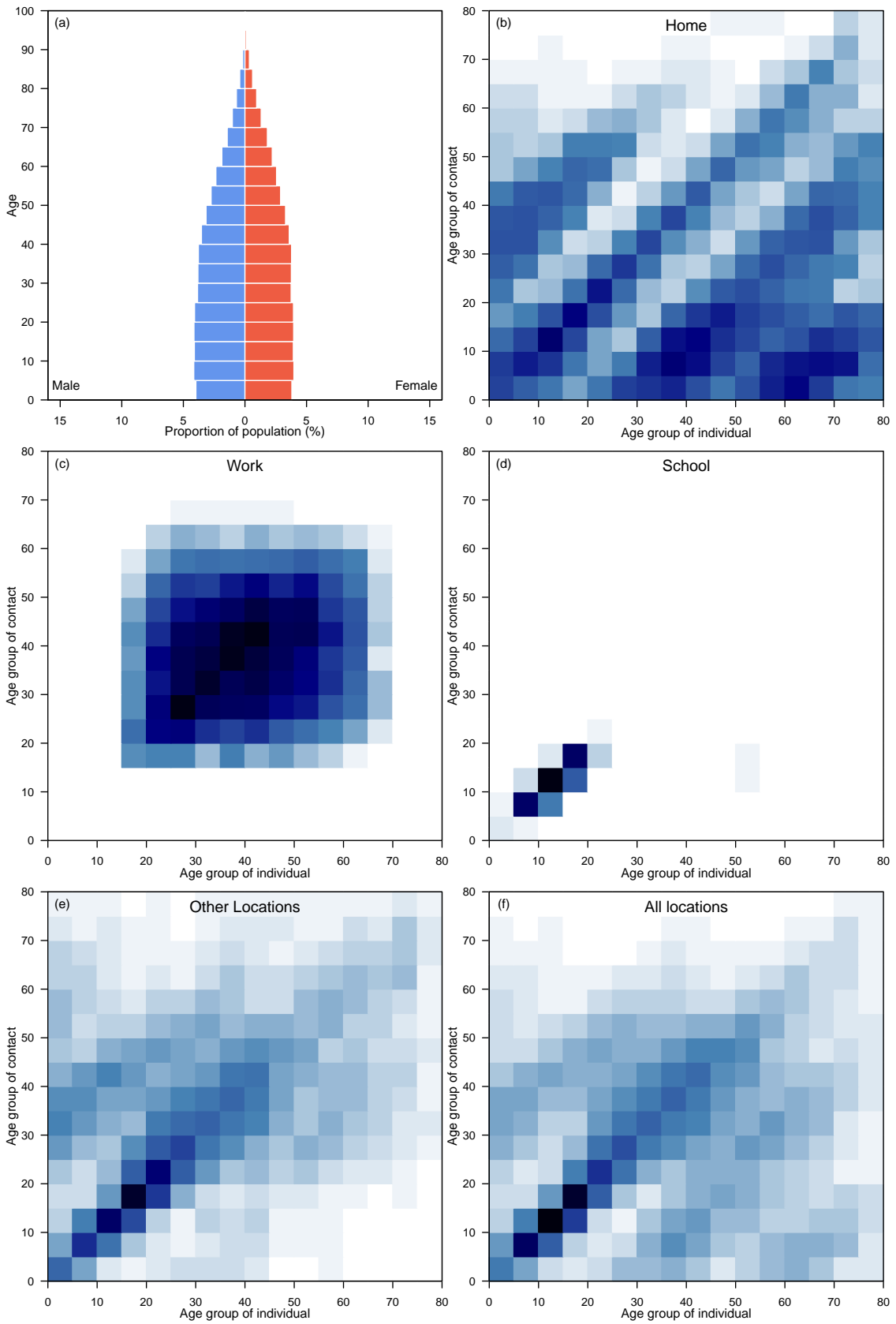
Turkmenistan
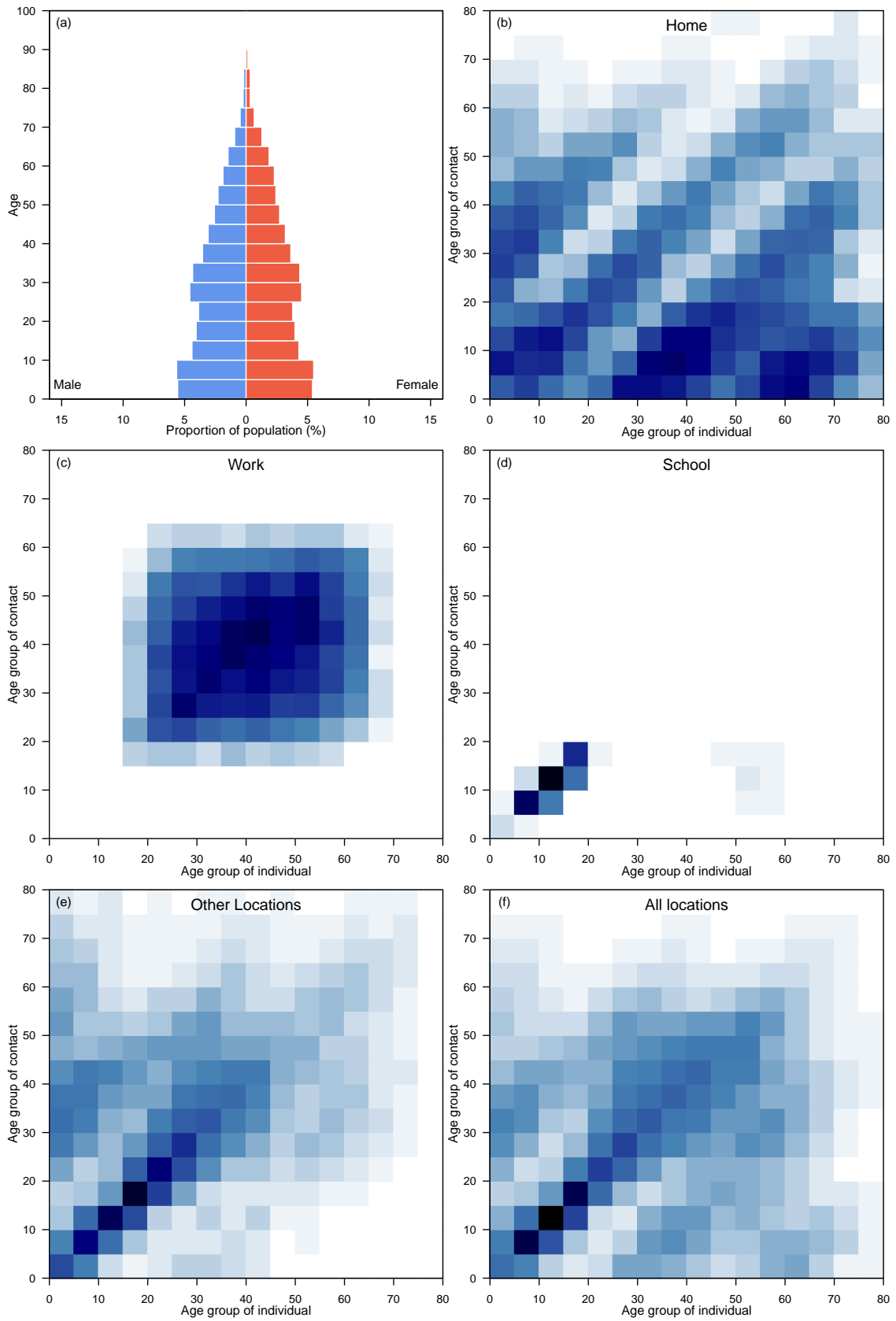
Uganda
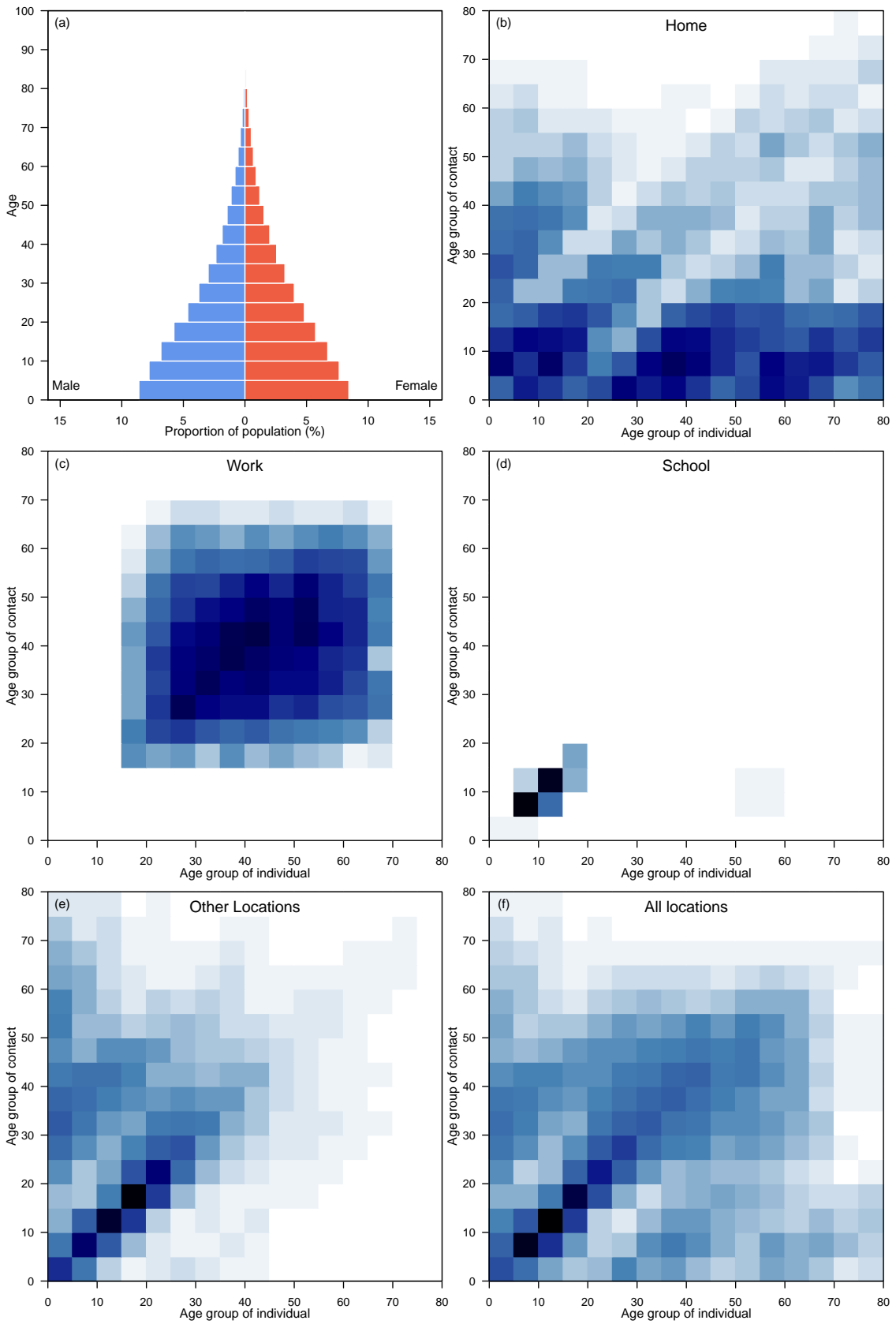
Ukraine
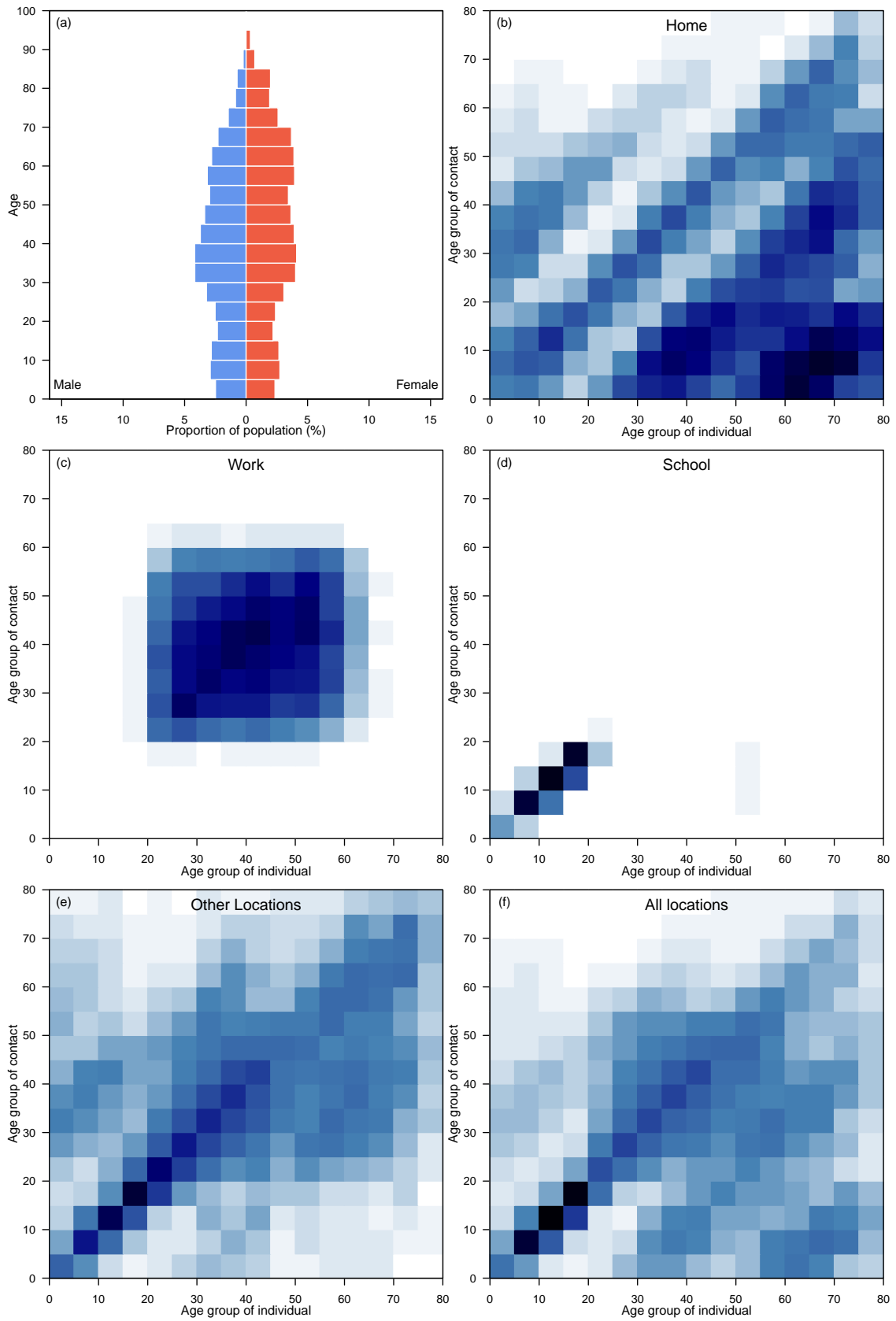
United Arab Emirates
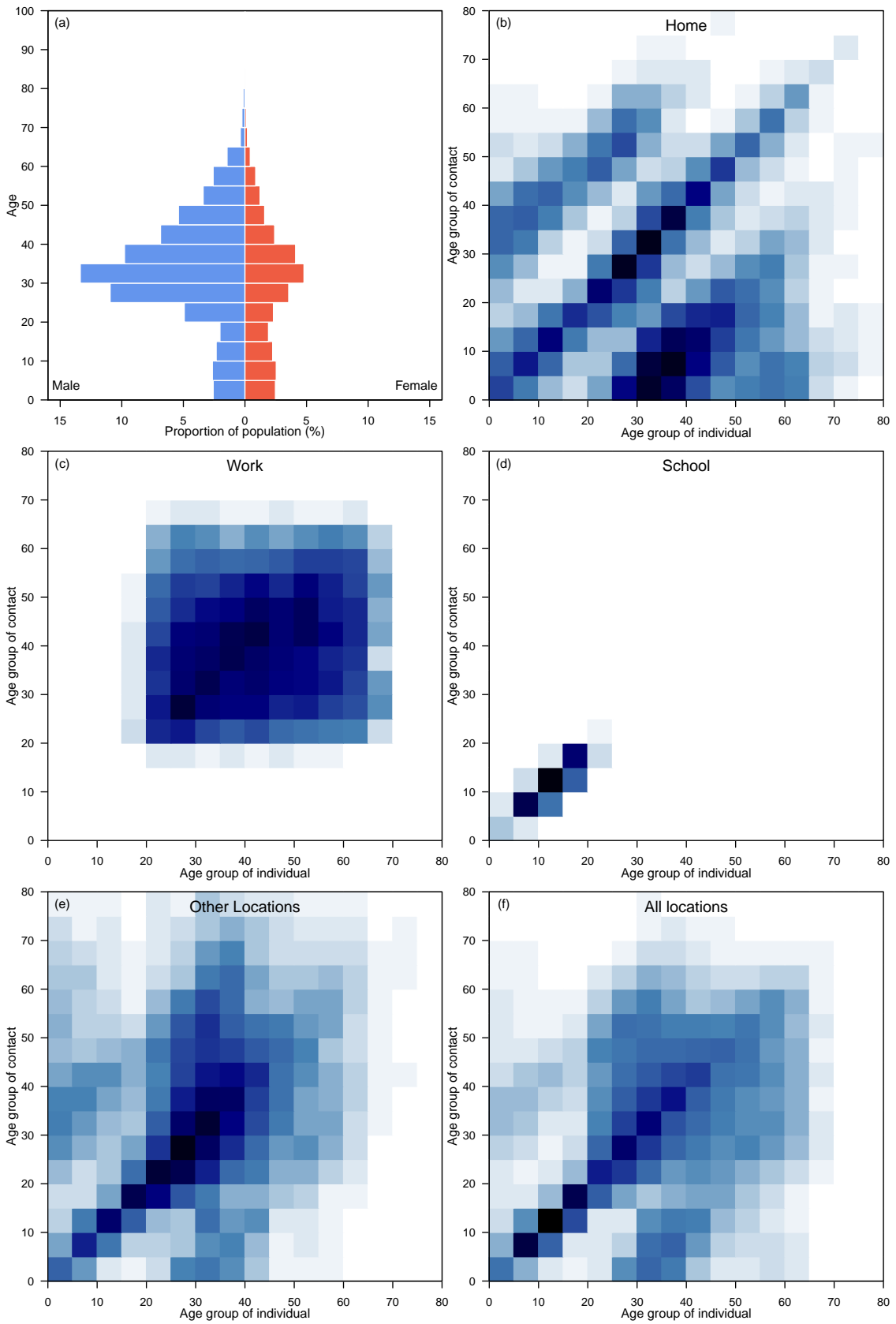
United Kingdom
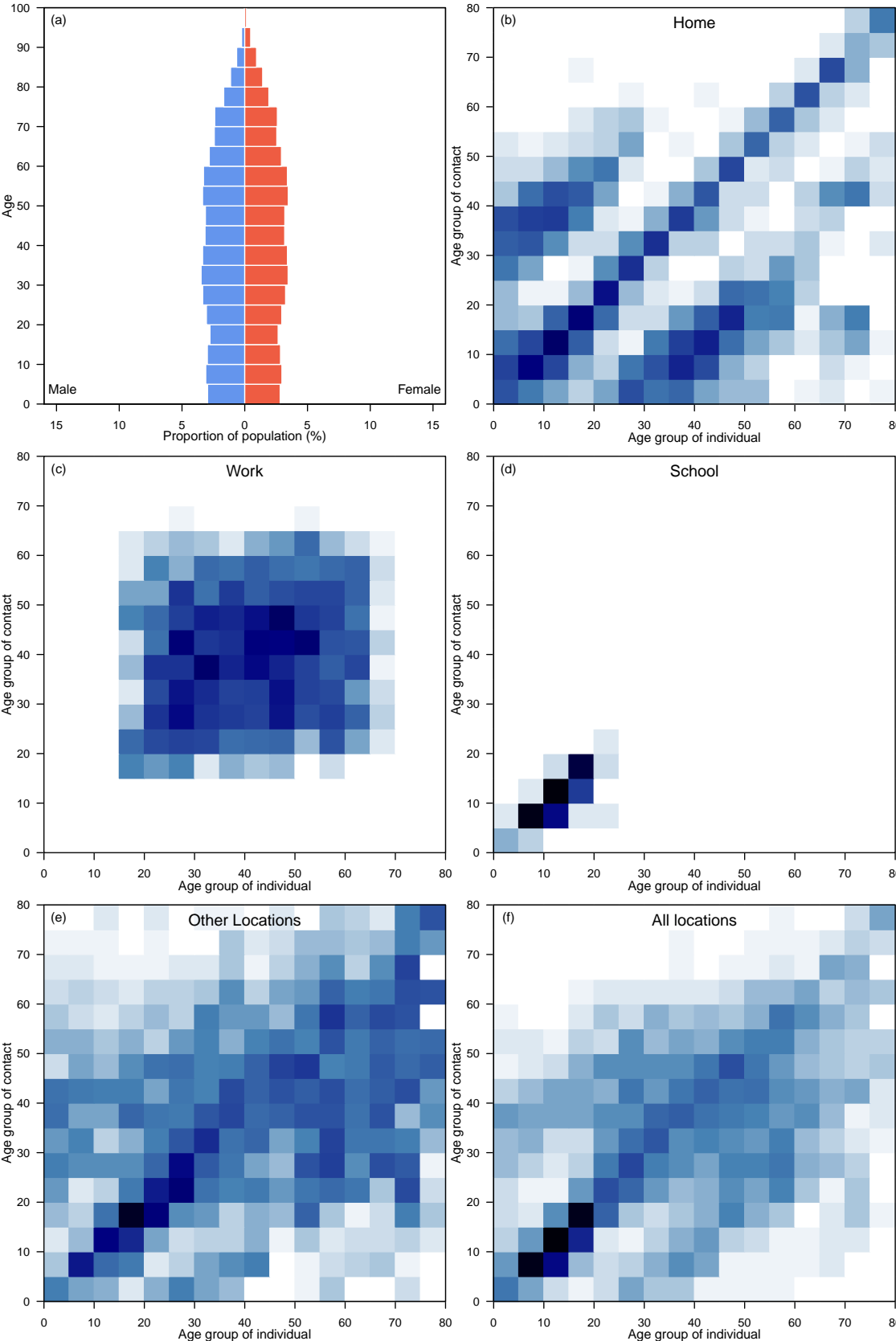
United Republic of Tanzania
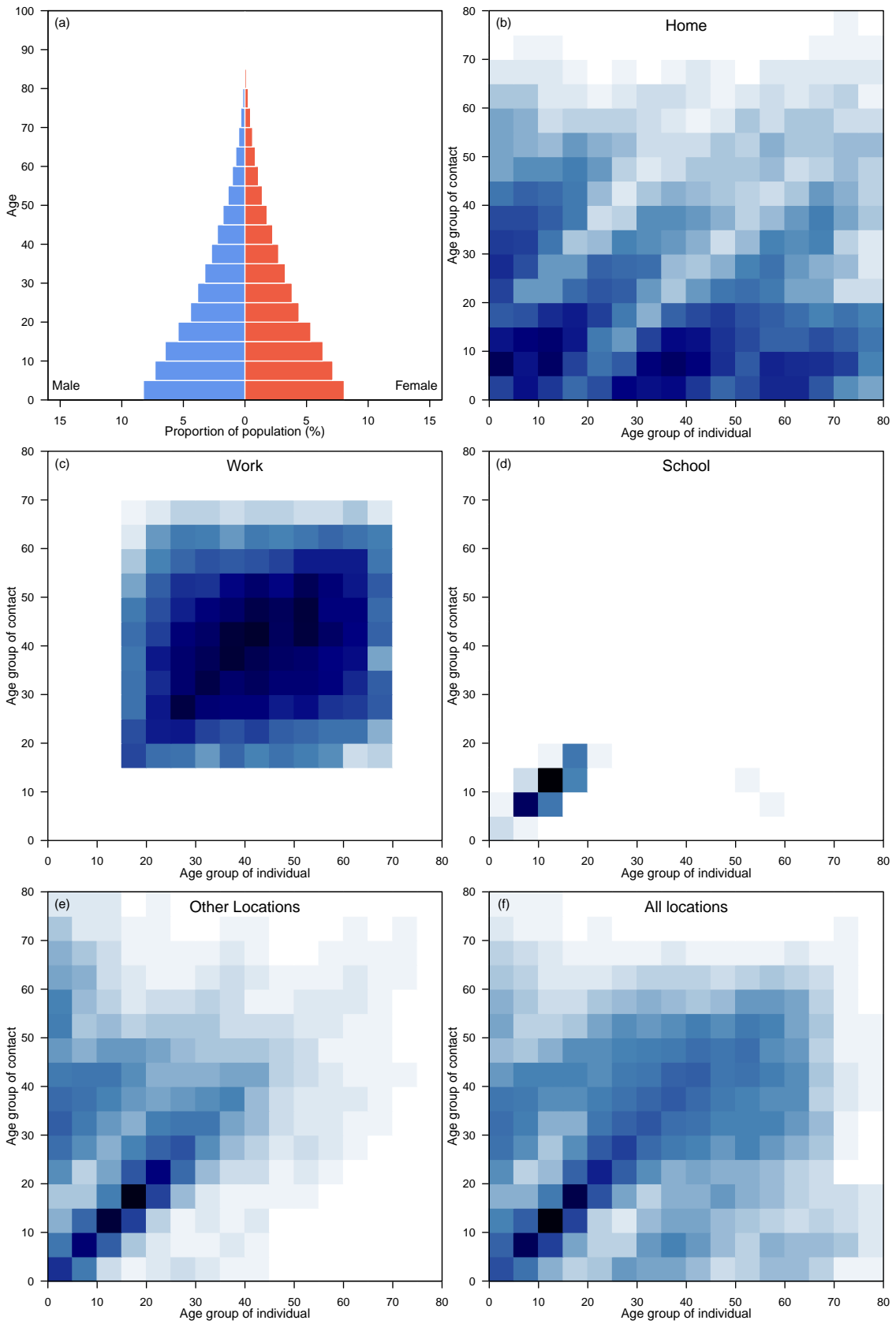
United States of America

Uruguay

Uzbekistan

(a) Male Female

(b) Home

(c) Work

(d) School

(e) Other Locations

(f) All locations
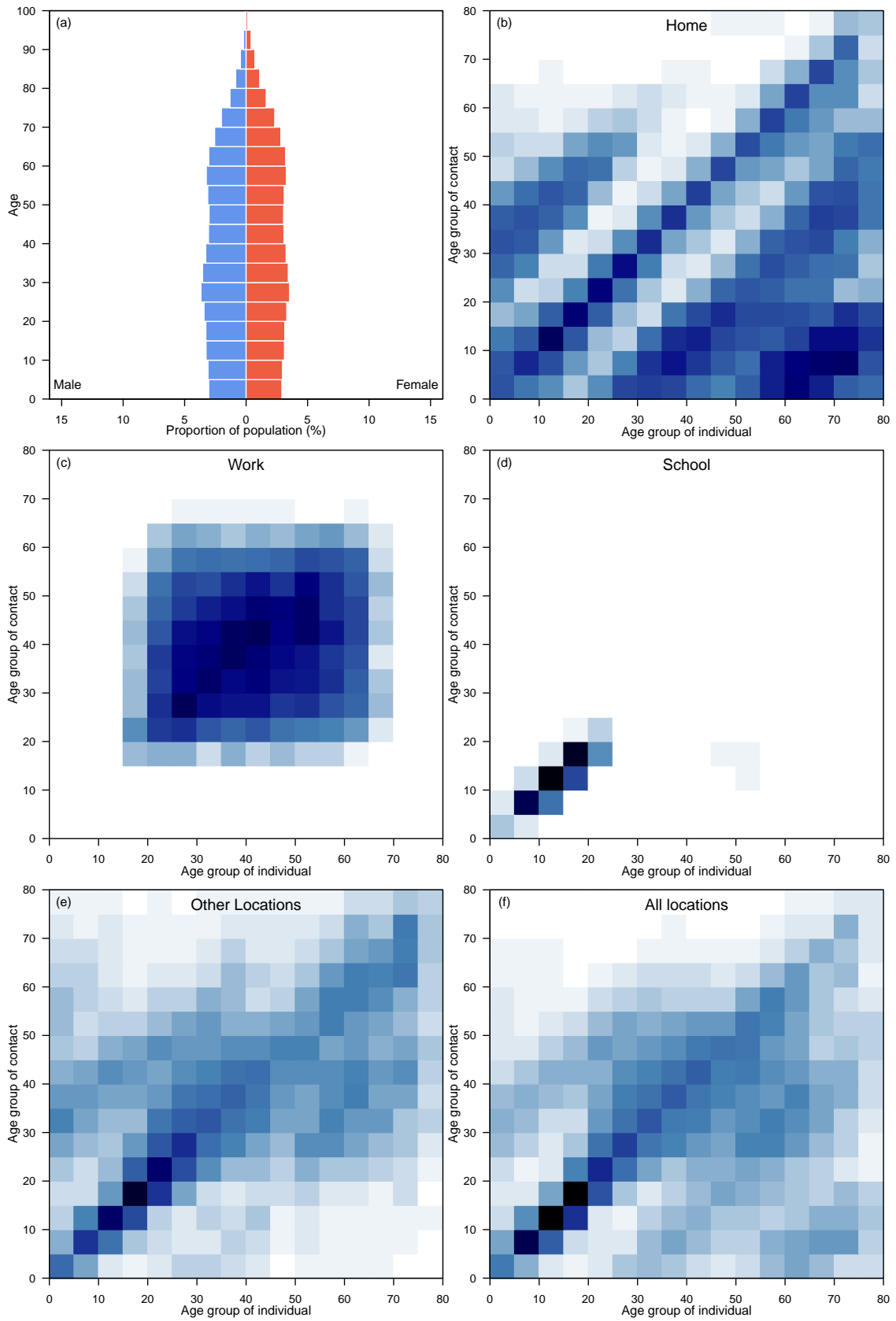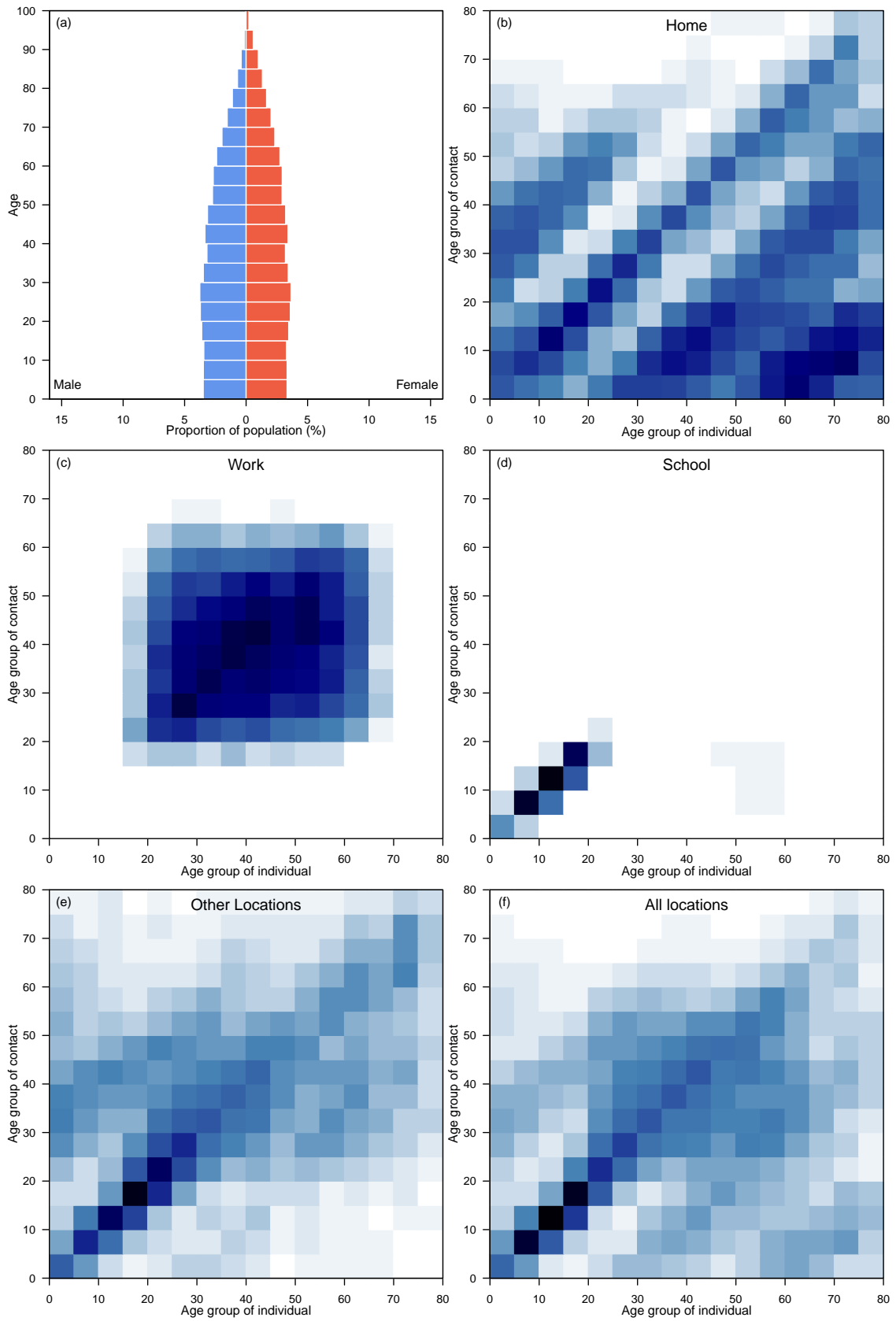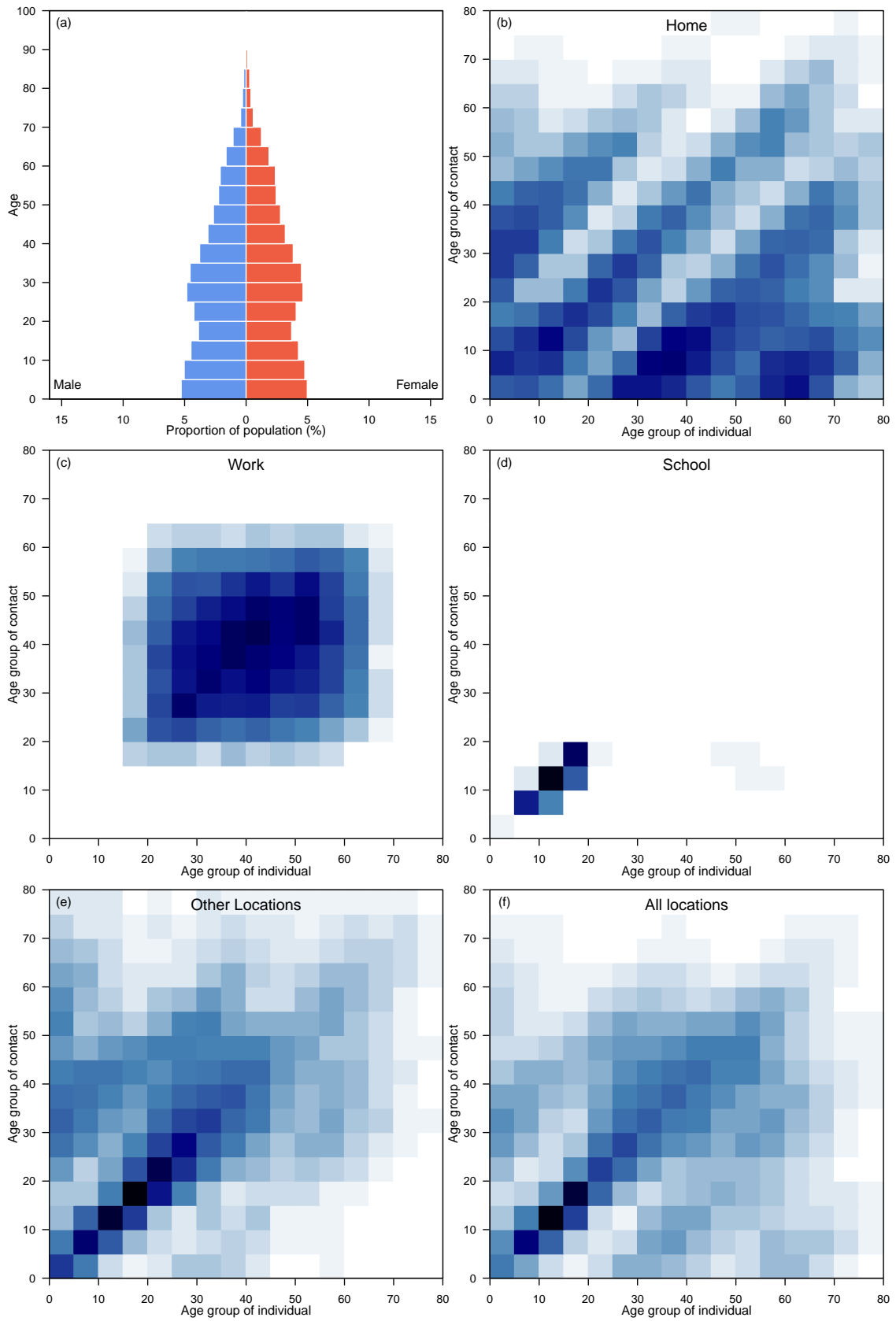
Vanuatu
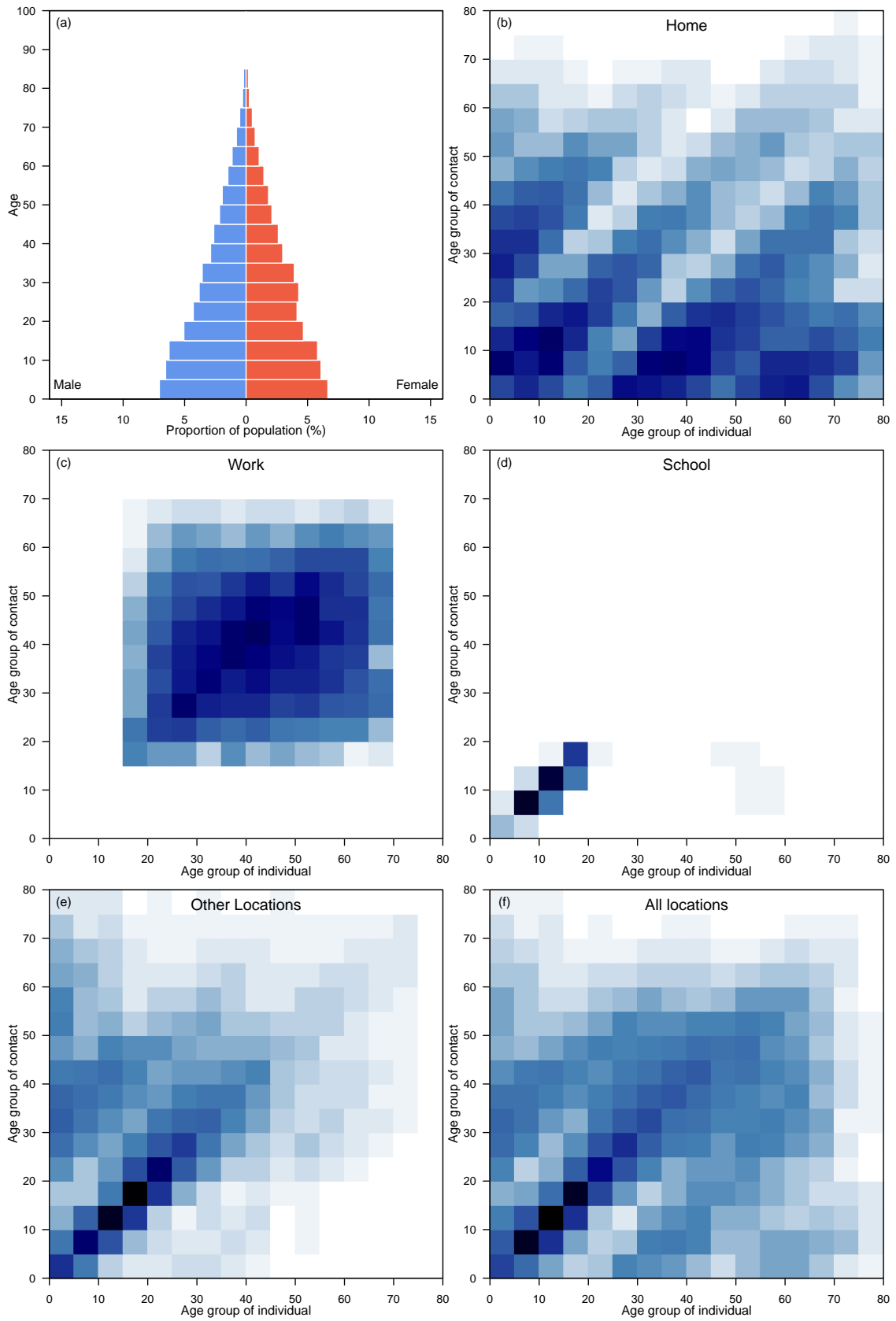
Venezuela (Bolivarian Republic of)
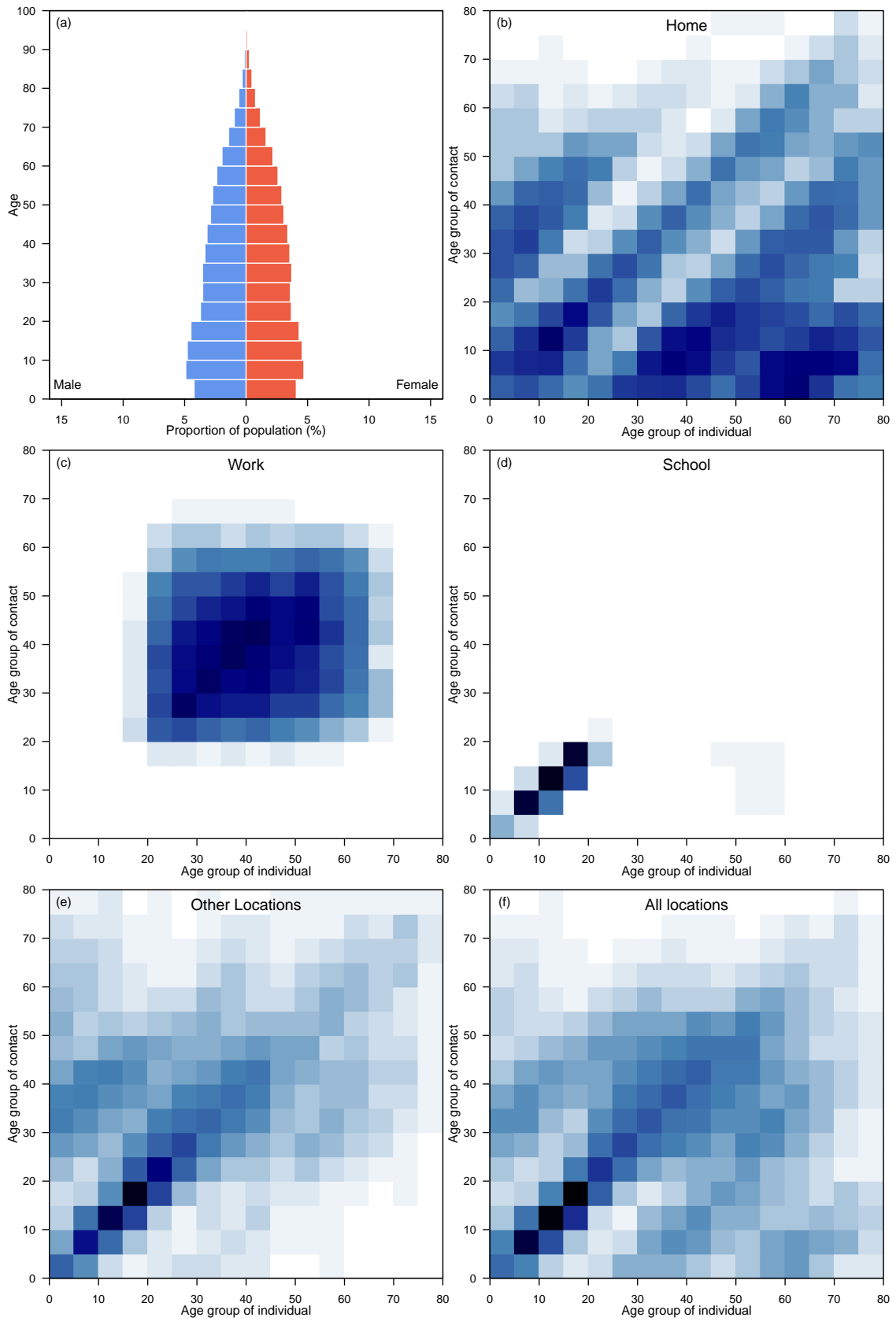
Viet Nam
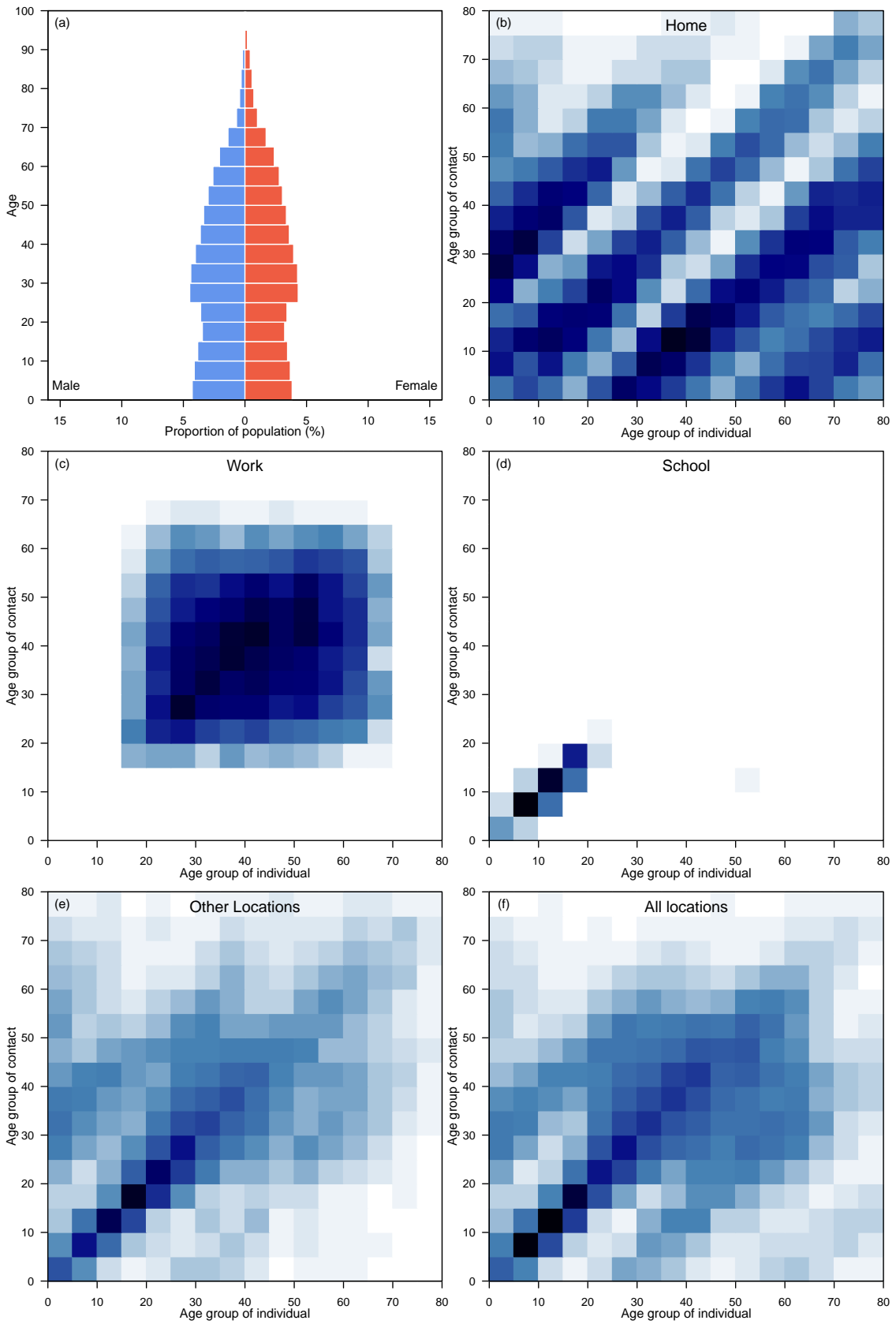
Yemen
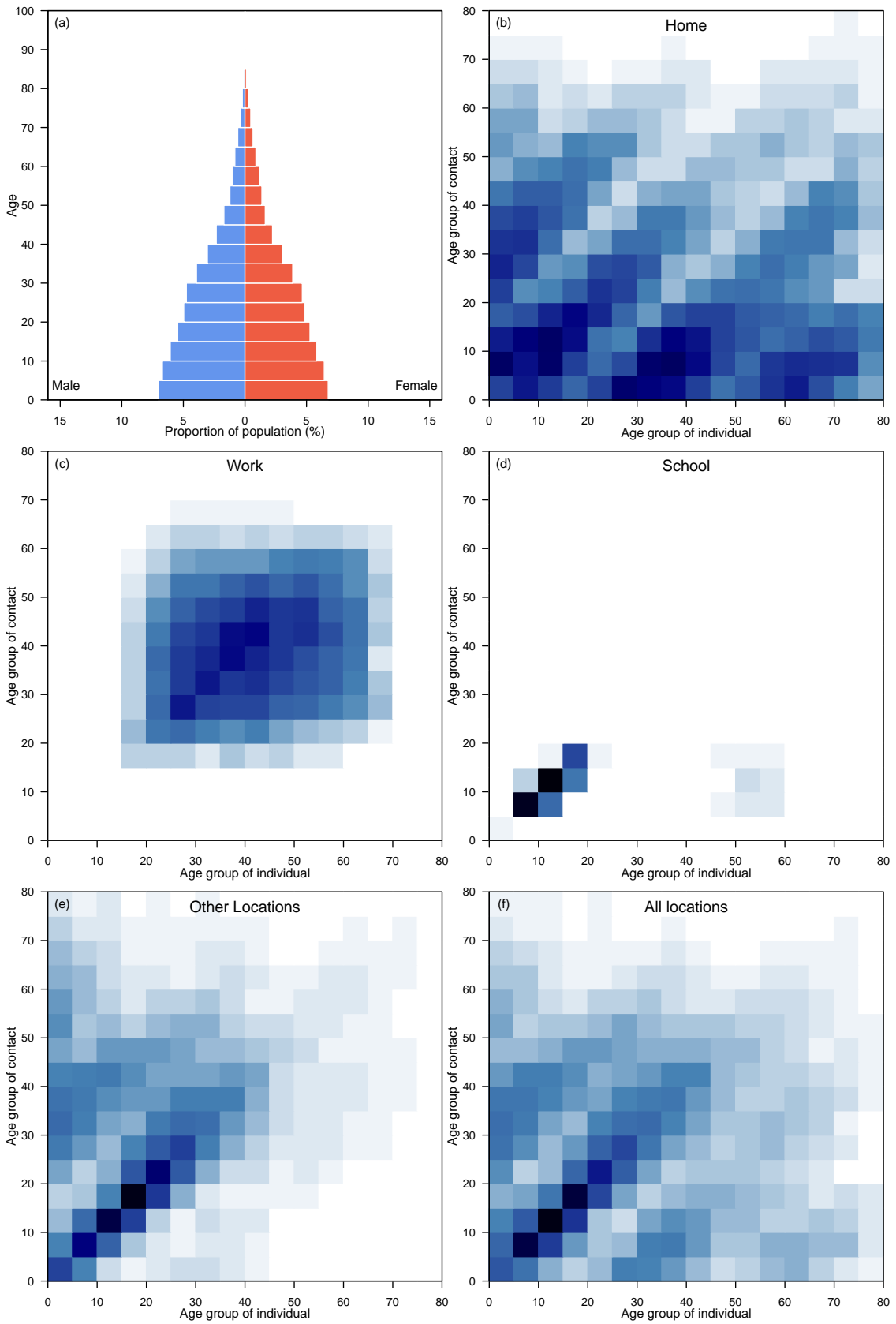
Zambia
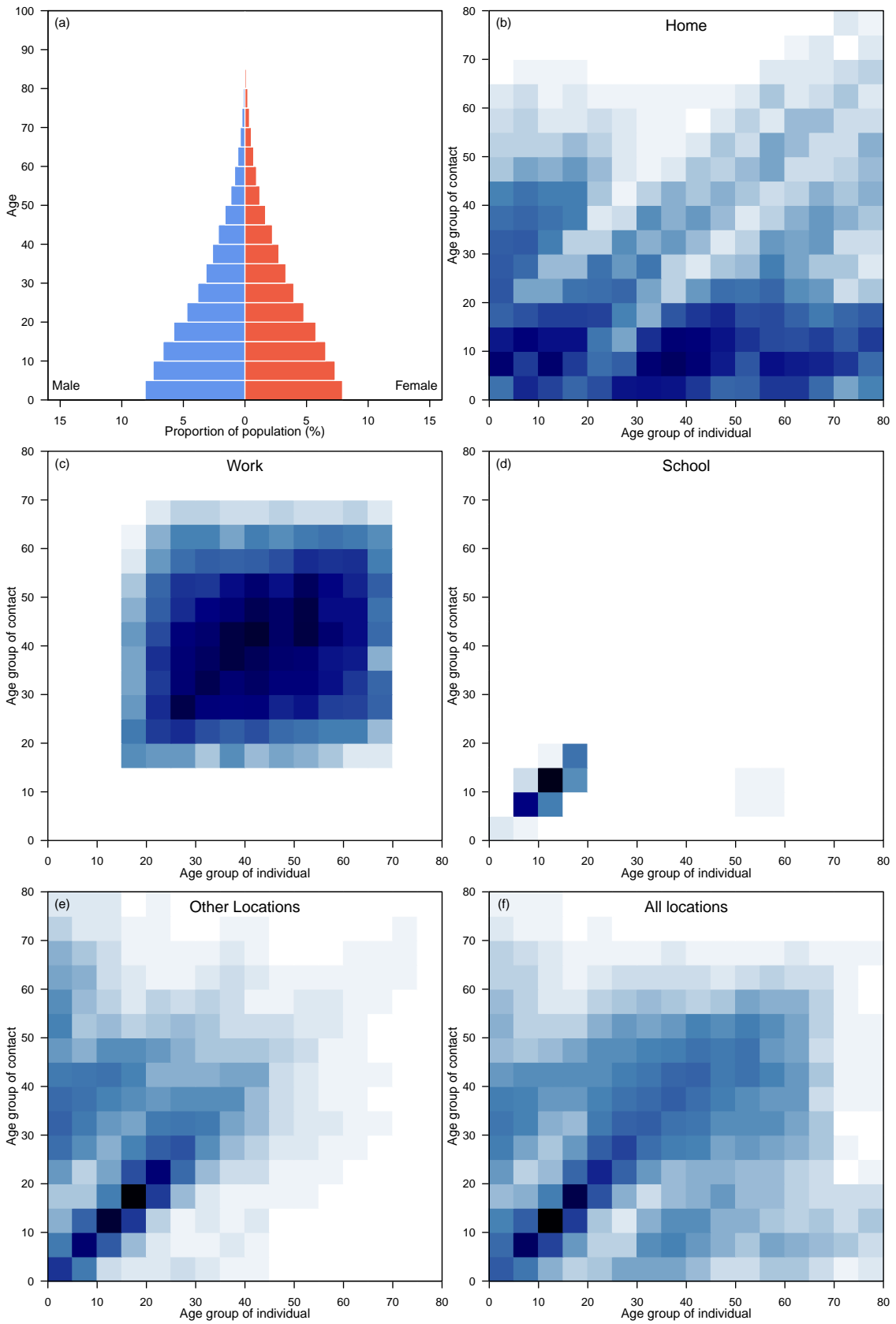
Zimbabwe
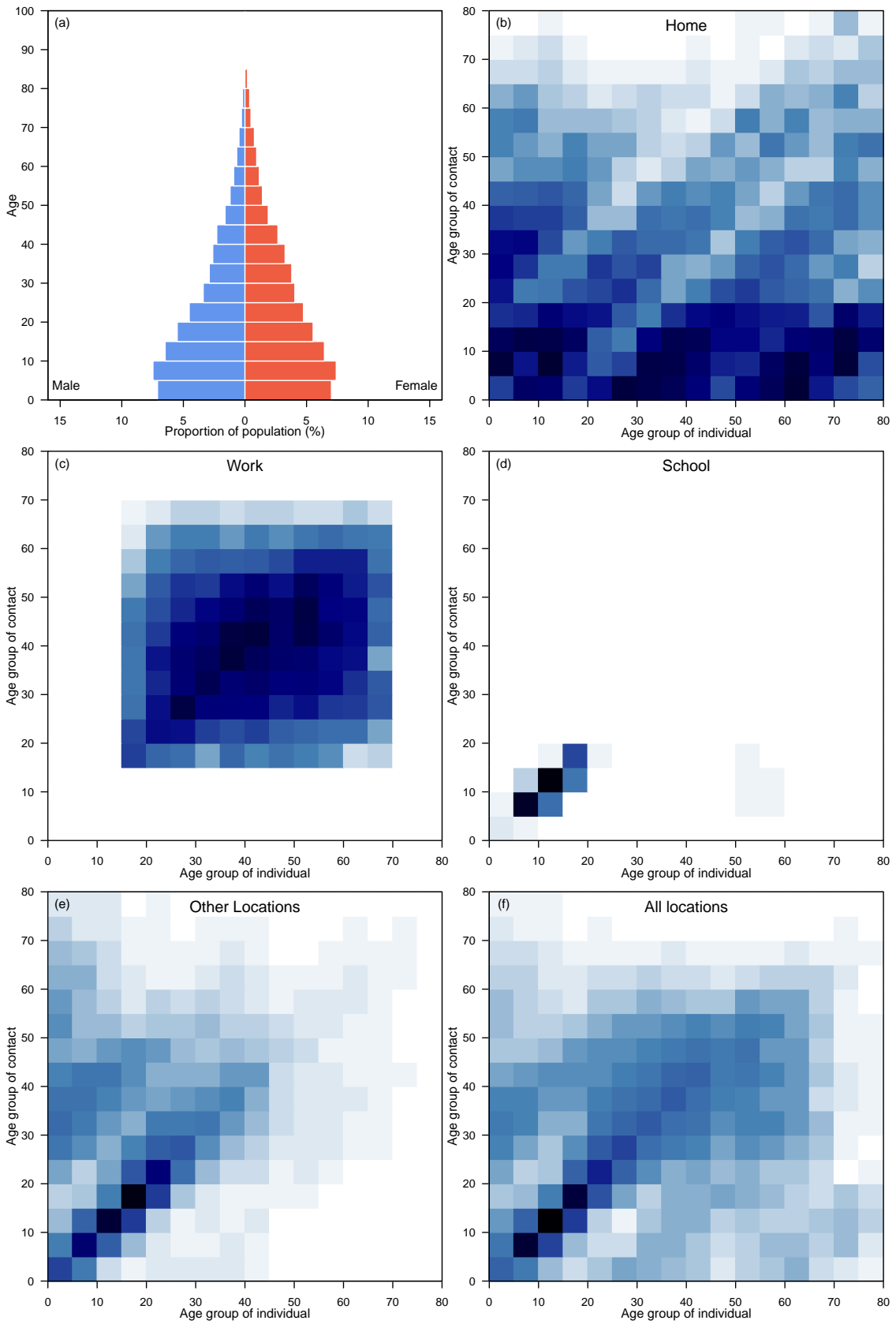
2.4 Comparison of the empirical and 2017 synthetic age-specific contact matrices in ten geographical regions

The 2017 synthetic matrices provide validated approximations to age-and-location-specific contact matrices for 152 geographical regions. We compare the normalised empirical matrices with to the normalised 2017 synthetic matrices.

Empirical contact survey
normalised

Synthetic 2017
normalised

China

France

China, Hong Kong SAR

Kenya

Peru

Age of contact

Age of individual

Nomalised mean number of contacts

Empirical contact survey
normalised

Synthetic 2017
normalised

Russian Federation — r = 0.91

South Africa — r = 0.9

Uganda — r = 0.85

Viet Nam — r = 0.85

Zimbabwe — r = 0.63

Age of individual

Nomalised mean number of contacts
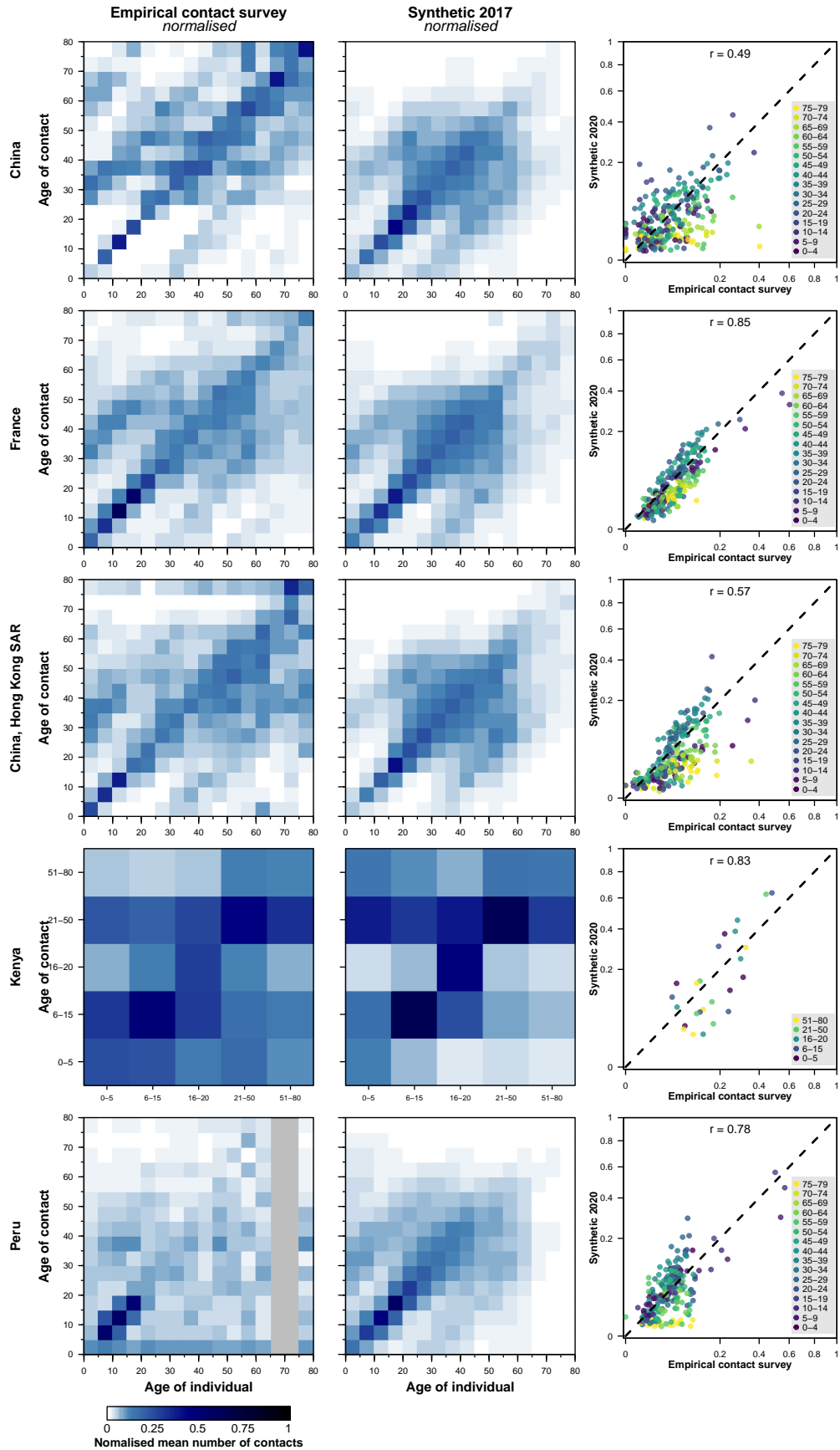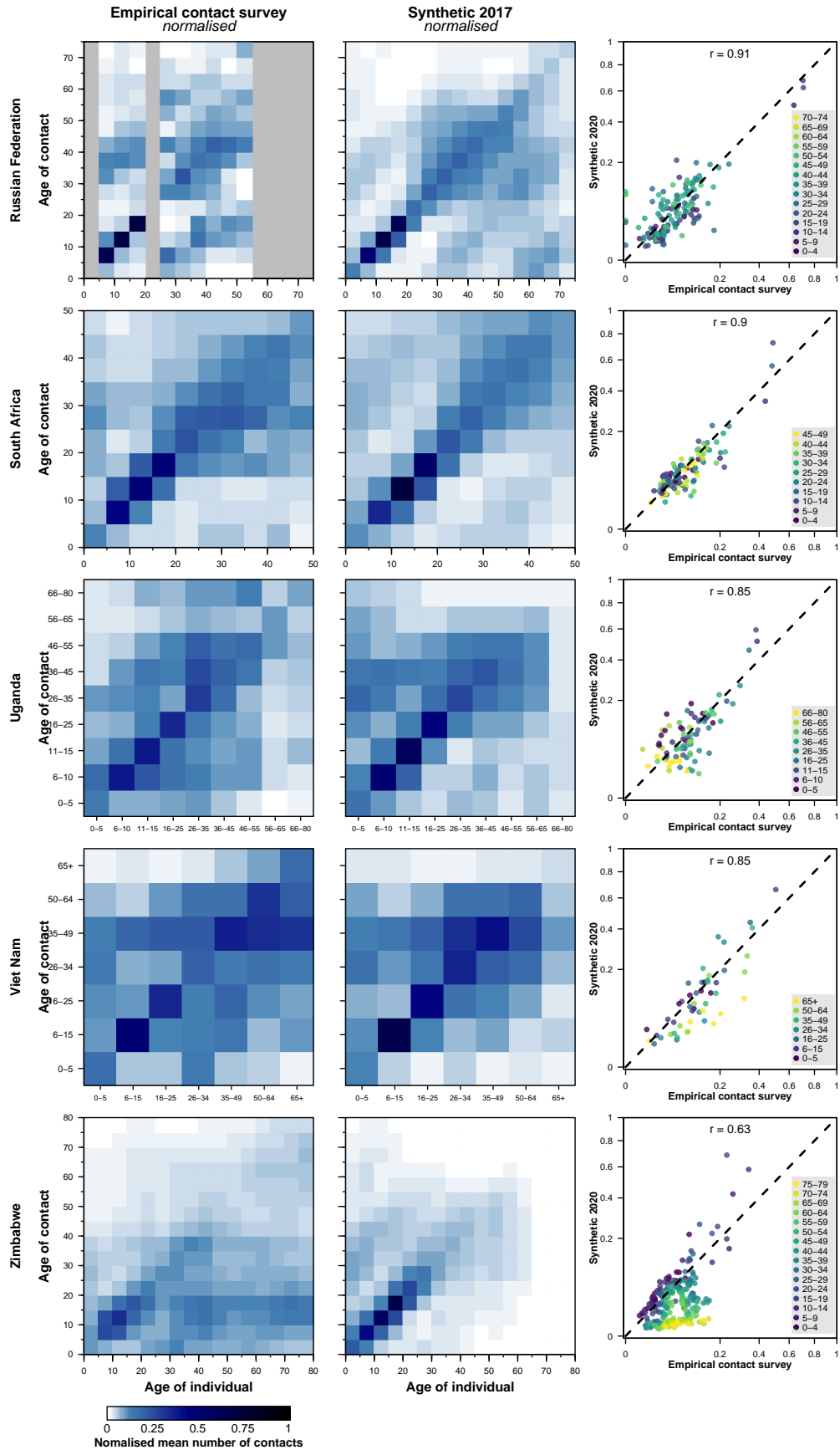
2.5 Age-specifc infection attack rate of COVID-19

The age-specific infection attack rate for the unmitigated epidemic under different contact matrices is shown in the boxplots with boxes bounded by the interquartile range (25th and 75th percentiles), the median in white and, whiskers spanning the $2.5 - 97.5^{th}$ percentiles (Figure 2). Six contact matrices were considered in the COVID-19 modelling: the empirically-constructed contact matrices at the study-year and adjusted for the 2020 population, the 2017 synthetic matrices, and the updated synthetic matrices at the national, rural, or urban settings. Models parameterised with the empirical and synthetic matrices generated similar findings with few differences observed in the age-specific infection attack rate (Figure 2) in age groups where the empirical matrices have missing or aggregated age groups (Table 4). This finding means that synthetic contact matrices may be used in modelling outbreaks in settings for which empirical studies have yet to be conducted.
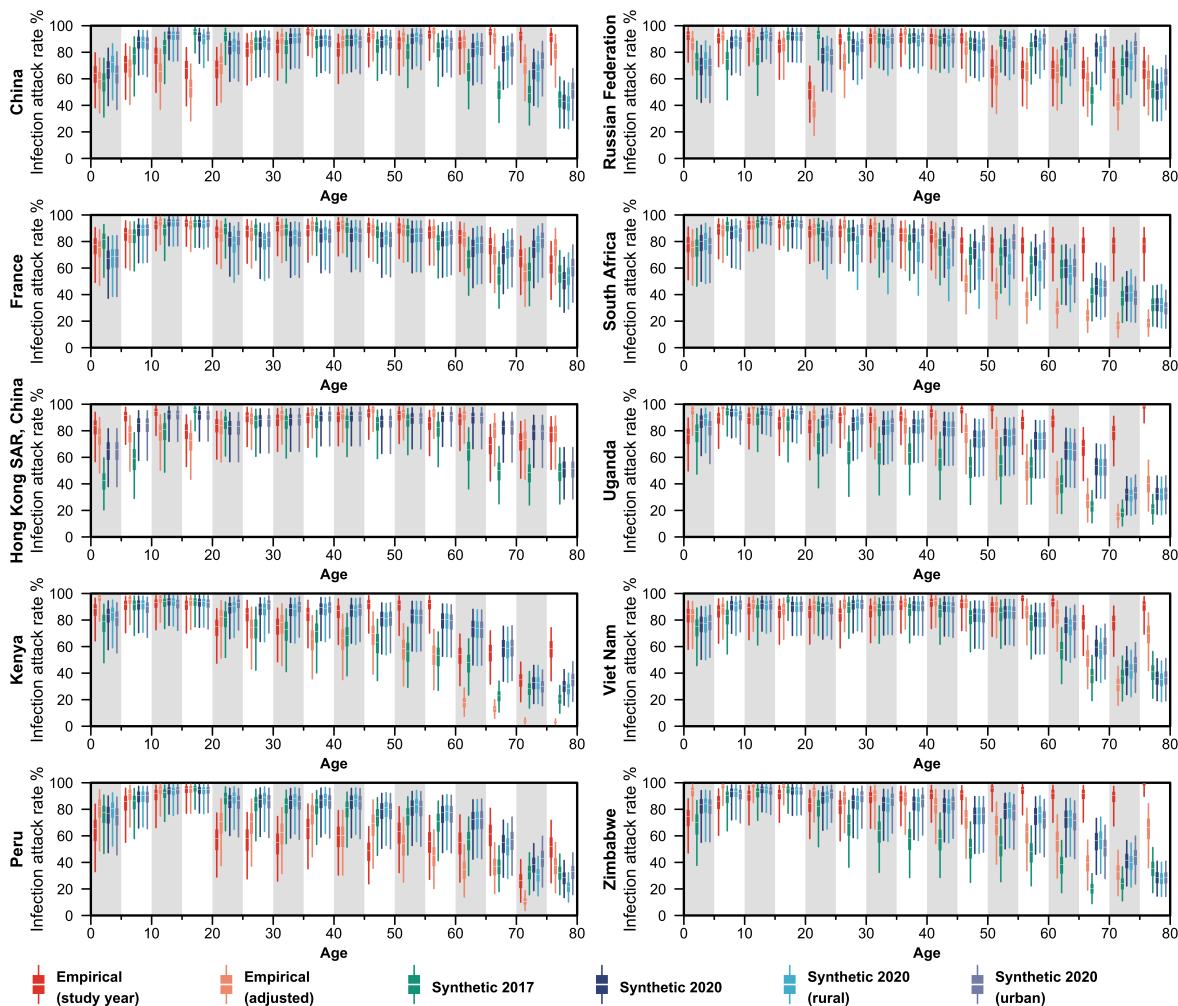
Figure 2: Age-specifc infection attack rate of COVID-19.

Table 4: Possible reason for discrepancies in age-specific infection attack rates

Discrepancies in age-specific infection attack rates	Survey-specific details	Possible reason for discrepancies
Underestimation of infection attack rate in the oldest age group (i.e., 75+ year-olds) in China, and Hong Kong SAR, China	Synthetic matrices report a lower proportion of contacts in other locations for individuals 75+ year-olds compared to empirical matrices in Shanghai and Hong Kong SAR	Individuals 75+ year old in two highly densely populated cities in China appear to report more contacts outside home and work than their counterparts in the POLYMOD surveyed countries in Europe. In the HK SAR survey, this may be due to a much larger proportion of older individuals choosing to use paper rather than online questionnaires; the survey found that individuals using paper surveys reported much higher numbers of contacts than those using online questionnaires.
Overestimation of infection attack rate in the 10–20-year-olds in China	Synthetic matrices report a higher proportion of contacts in other locations for individuals 10–20-year-olds compared to empirical matrices in Shanghai, China	Individuals 10–20-year-olds in Shanghai, China appear to report fewer contacts at other locations than their counterparts in the POLYMOD surveyed countries in Europe.
Differences in infection attack rate in individuals 50+ year-olds in Kenya	In the study in Kilifi, Kenya, individuals aged 21–50 were aggregated and the oldest age group in the survey is 51+ year-olds.	These differences are observed in age groups where the empirical matrices have aggregated age groups, which may obscure large differences within ages in that aggregated age group.
Differences in infection attack rate in individuals aged 70–74-year-olds in Peru	The study in San Marcos, Cajamarca-Peru did not survey any individuals aged 70–74-year-old.	These differences are observed in age groups where the empirical matrices have missing age groups. Contacts in 70-74 year olds in the empirical matrices are based on assuming reciprocity of contacts for younger people reporting contacts with this age group, which we know from other surveys may lead to incorrect estimations.

Discrepancies in age-specific infection attack rates	Survey-specific details	Possible reason for discrepancies
Differences in infection attack rate in individuals aged 20–24-year-olds in Russia	The study in Tomsk, Russia did not survey any individuals aged 20–24-year-old.	These differences are observed in age groups where the empirical matrices have missing age groups. Contacts in 20-24 year olds in the empirical matrices are based on assuming reciprocity of contacts for other people reporting contacts with this age group, which we know from other surveys may lead to incorrect estimations.
Differences in infection attack rate in individuals 50+ year-olds in South Africa	The oldest age group in the survey conducted in a South African township is 45+ year-olds. After adjusting for the South African 2020 population age composition, the differences narrow.	These differences are observed in age groups where the empirical matrices have aggregated age groups, which may obscure large differences within ages in that aggregated age group.
Differences in infection attack rate in individuals 65+ year-olds in Uganda	The oldest age group in the survey conducted in a Sheema North Sub-District, Uganda is 65+ year-olds. These differences narrow after adjusting for Uganda's 2020 population age composition.	These differences are observed in age groups where the empirical matrices have aggregated age groups, which may obscure large differences within ages in that aggregated age group.
Differences in infection attack rate in individuals 65+ year-olds in Vietnam	In the study in Red River Delta, North Vietnam, individuals aged 50–64 were aggregated and the oldest age group in the survey is 65+ year-olds. After adjusting for the Vietnamese 2020 population age composition, the differences narrow.	These differences are observed in age groups where the empirical matrices have aggregated age groups, which may obscure large differences within ages in that aggregated age group.
Differences in infection attack rate in individuals 65+ year-olds in Zimbabwe	Compared to the empirical matrices in Manicaland, Zimbabwe, the synthetic matrices underestimate the number of contacts for individuals 65+ year-olds at other locations. The differences narrow after adjusting for the 2020 population age composition in Zimbabwe. The survey was administered during the school holiday in one of the two sites; they found a lower number of contact was reported in schools. As school-going children are not in school, they could be spending more time at home and possibly with their grandparents; and this could be one possible reason for the high contacts in older individuals with the younger population.	Synthetic matrices report fewer contacts in other locations for individuals aged 65+ years compared to the empirical matrices in Manicaland, Zimbabwe. The differences narrow after adjusting for the 2020 population age composition in Zimbabwe. The survey was administered during the school holiday in one of the two sites; they found a lower number of contact was reported in schools. As school-going children are not in school, they could be spending more time at home and possibly with their grandparents; and this could be one possible reason for the high contacts in older individuals with the younger population.

References

1. United Nations Department of Economic and Social Affairs Population Division. World Population Prospects. 2019. Available: https://population.un.org/wpp/

2. United Nations Department of Economic and Social Affairs Population Division. Urban and Rural Population by Age and Sex, 1980-2015. Available: https://www.un.org/en/development/desa/population/publications/dataset/urban/urbanAndRuralPopulationByAgeAndSex.asp

3. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. Riley S, editor. PLoS Medicine. 2008;5: e74. doi:10.1371/journal.pmed.0050074

4. Demographic and Health Surveys. The DHS Program - Demographic and Health Survey (DHS). Available: https://dhsprogram.com/What-We-Do/Survey-Types/DHS.cfm

5. Prem K, Cook AR, Jit M. Projecting social contact matrices in 152 countries using contact surveys and demographic data. PLoS Computational Biology. 2017;13: e1005697. doi:10.1371/journal.pcbi.1005697

6. UNESCO Institute for Statistics. UIS Statistics. Available: http://data.uis.unesco.org/

7. International Labour Organization. Labour force by sex and age.

8. OECD. Teachers by age (indicator). 2020. doi:10.1787/93af1f9d-en

9. International Labour Organization. Labour force participation rate by sex, age and rural / urban areas.

10. OECD. Differences in rural and urban schools'; student-teacher ratio and class size, 2015. OECD Publishing; 2018.

11. Davies NG, Kucharski AJ, Eggo RM, Gimma A, Edmunds WJ, Jombart T, et al. Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. The Lancet Public Health. 2020;0. doi:10.1016/s2468-2667(20)30133-x

12. van Zandvoort K, Jarvis CI, Pearson C, Davies NG, CMMID COVID-19 working Group, Russell TW, et al. Response strategies for COVID-19 epidemics in African settings: a mathematical modelling study. medRxiv. 2020; 2020.04.27.20081711. doi:10.1101/2020.04.27.20081711

13. Jarvis CI, Van Zandvoort K, Gimma A, Prem K, Klepac P, Rubin GJ, et al. Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK. BMC Medicine. 2020;18: 124. doi:10.1186/s12916-020-01597-8

14. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. New England Journal of Medicine. 2020. doi:10.1056/nejmoa2001316

15. Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, et al. Epidemiology and Transmission of COVID-19 in Shenzhen China: Analysis of 391 cases and 1,286 of their close contacts. medRxiv. 2020; 2020.03.03.20028423. doi:10.1101/2020.03.03.20028423

16. Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. International Journal of Infectious Diseases. 2020;93: 284–286. doi:10.1016/j.ijid.2020.02.060

17. Liu Y, Funk S, Flasche S. The contribution of pre-symptomatic infection to the transmission dynamics of COVID-2019. Wellcome Open Research. 2020;5: 58. doi:10.12688/wellcomeopenres.15788.1

18. Davies NG, Klepac P, Liu Y, Prem K, Jit M, Eggo RM. Age-dependent effects in the transmission and control of COVID-19 epidemics. Nature Medicine. 2020; 1–7. doi:10.1038/s41591-020-0962-9

19. Linton NM, Kobayashi T, Yang Y, Hayashi K, Akhmetzhanov AR, Jung S-m, et al. Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data. Journal of Clinical Medicine. 2020;9: 538. doi:10.3390/jcm9020538

20. Cao B, Wang Y, Wen D, Liu W, Wang J, Fan G, et al. A Trial of Lopinavir–Ritonavir in Adults Hospitalized with Severe Covid-19. New England Journal of Medicine. 2020;382: 1787–1799. doi:10.1056/NEJMoa2001282

21. The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. The epidemiological characteristics of an outbreak of 2019 novel coronavirus disease (COVID-19)—China. 2020. China CDC Wkly. 2020;2: 113–122.