Foundations of Data Science

Capstone Project

World Population Projections

by Rabeyl Aslam

14 October 2016

Introduction

Population projections are routinely used by private, public and non-profit organizations for policy development (“Uses and Limitations”, 2016). For example, projections are used in:

- Estimating central and local finance allocation
- Planning housing and land use
- Health care planning: for modelling and projecting health care indicators
- Weighting national and regional surveys
- Creating teacher workforce models at a national and local level
- Assessing ageing population and understanding its implications

Therefore, accurate and high-quality population projections are imperative for good policy-making. **For this assignment, we will be estimating aggregate, age-specific, and sex- specific population projections until 2100 for 200+ countries.**

Fertility, life expectancy and migration statistics are the three main independent variables for estimating population projections (Keneda & Bremner, 2014). Fertility is expressed as total fertility rate (TFR). Total fertility rate is the average number of children that a woman will have in her lifetime if the rate of childbearing in a given year remains unchanged over her entire life (Keneda & Bremner, 2014). Life expectancy at birth is the average number of years a person is expected to live if (s)he were to pass through life subject to age-specific mortality rates of a given period (“Life Expectancy at”, 2000). For this study, we only take international migration into account. An international migrant is defined as a person who is living in a country other than his or her country of birth (Menozzi, 2015).

For this assignment, we obtain data from United Nations Population Division Database (data points from 1950 to 2012). This data is readily available in R software in 'wpp2012' package. This package contains estimates of total fertility rates for each country, sex-specific rates of life expectancy for each country, age-specific and sex-specific estimates for population for each country, age-specific and sex-specific mortality rates for each country, age-specific and sex-specific net migration for each country, sex ratios at birth (as a ratio of female to male) for each country, distribution of age-specific fertility rates for each country, and the location dataset.

In order to estimate population projections for our assignment the following data input is required for each country:

- Sex-specific and age-specific population estimates at the initial time $t = 0$
- Projections of total fertility rate (TFR)
- Projections of fertility distribution over ages
- Projections of sex ratio at birth
- Projections of male and female life expectancy at birth (e0)
- Historical data on sex- and age-specific death rates (for $t \leq 0$)
- Projections of sex- and age-specific net migration

The input projections for fertility, life expectancy and migration mentioned above are generated using data from 'wpp2012' package in R software. After generating the required data input we estimate population projections for each country. The next section will describe this methodology in more detail.

Methodology

As per Sevcikova et al. (2013), population projections are estimated using the following demographic balancing equation:

$$\mathbf{P_{c,t} = P_{c,t-1} + B_{c,t} - D_{c,t} + M_{c,t}} \text{ ———————————————————— } \mathbf{(i)}$$

where $P_{c,t}$ is the population prediction in country c at time t,

$P_{c,t-1}$ is the population prediction in country c at time t−1,

$B_{c,t}$ is the number of births in country c at time t,

$D_{c,t}$ the number of deaths in country c at time t,

$M_{c,t}$ net migration in country c at time t.

Theoretically, this equation implies that for any given time period the population of a country is dependant upon population prediction of that country in the previous time period, number of births in the country at that time, number of deaths in the country at that time and the net migration in country at that time.

Our analysis is divided into three main steps. The first two steps are the feature engineering part of the analysis — we estimate fertility and life expectancy parameters for each country. The final step is machine learning part of the analysis — we estimate population projections for each country using the parameters we engineered in the previous two steps. The next section gives a detailed description of steps in R software — code is included.

R Analysis

As mentioned in the previous section, here are the three main steps of our analysis:

Feature engineering part

Step 1 — Simulating a large set of trajectories for future values of total fertility rates using Bayesian hierarchical estimates

- Done via ‘BayesTFR’ package in R software — note that these simulations take several hours.

R code for Step 1:

```
## Set working directory

tfr.dir <- "/Users/rabeylaslam/Desktop/CapstoneDataTFR"

## Estimate parameters of the Phase II model using Markov Chain Monte Carlo (MCMC)

mc1 <- run.tfr.mcmc(output.dir=tfr.dir, iter='10000', wpp.year=2012,
start.year=1950, present.year=2010, parellal = TRUE)

## Estimate the parameters of the Phase III model by MCMC

mc2 <- run.tfr3.mcmc(sim.dir=tfr.dir, iter='10000')

## Using the estimated parameters, generate future TFR trajectories

tfr.pred <- tfr.predict(sim.dir=tfr.dir, end.year=2100, nr.traj=1000, burnin=4000,
burnin3=4000)
```

These total fertility rate trajectories are used in Step 3 of our analysis.

Step 2 – Simulating a large and equal set of trajectories for future values of life expectancies using Bayesian hierarchical estimates

- Done via 'BayesLife' package in R software – note that these simulations take several hours

R code for Step 2:

```
## Set working directory

e0.dir <- "/Users/rabeylaslam/Desktop/CapstoneDataLifeExpectancy"

## Estimate parameters for female e0 (life expectancy) via MCMC

mc <- run.e0.mcmc(output.dir=e0.dir, sex="Female", iter='20000', wpp.year=2012,
parellal = TRUE)

## Using estimated parameters, generate future female and male e0 (life-
expectancy) trajectories

e0.pred <- e0.predict(sim.dir=e0.dir, end.year=2100, nr.traj=1000, burnin=8000)
```

These sex-specific life expectancy trajectories are used in Step 3 of our analysis.

Machine learning part

Step 3 — Converting trajectories from previous models to age-specific and sex- specific population projections using cohort component method

- Done via 'BayesPop' package in R software

R code for Step 3:

Set working directory

pop.dir <- "/Users/rabeylaslam/Desktop/CapstoneDataPopProj"

Predicting and storing population projections

pop.pred <- pop.predict(end.year=2100, start.year=1950, present.year=2010,

wpp.year=2012, output.dir=pop.dir, nr.traj=1000,

inputs=list(tfr.sim.dir=tfr.dir,

e0F.sim.dir=e0.dir, e0M.sim.dir="joint_"))

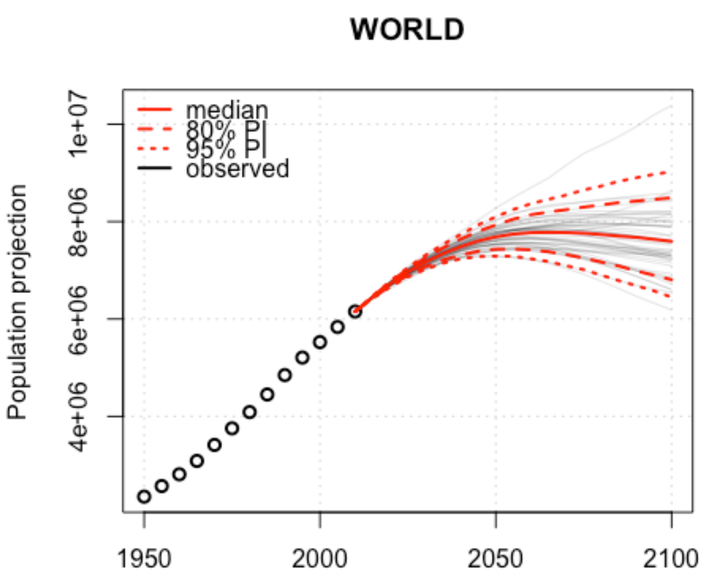
Note that we can access these population projections in a later session using separate R commands (entire R code for this analysis is provided separately). In the following sections results are shown and conclusions are drawn.

Results

From our analysis not only can we draw country-specific conclusions on population projections, but we can also ascertain regional and global trends by aggregating the results from individual countries.

Global Trends

The following two figures summarize world population projections (1950 to 2100).



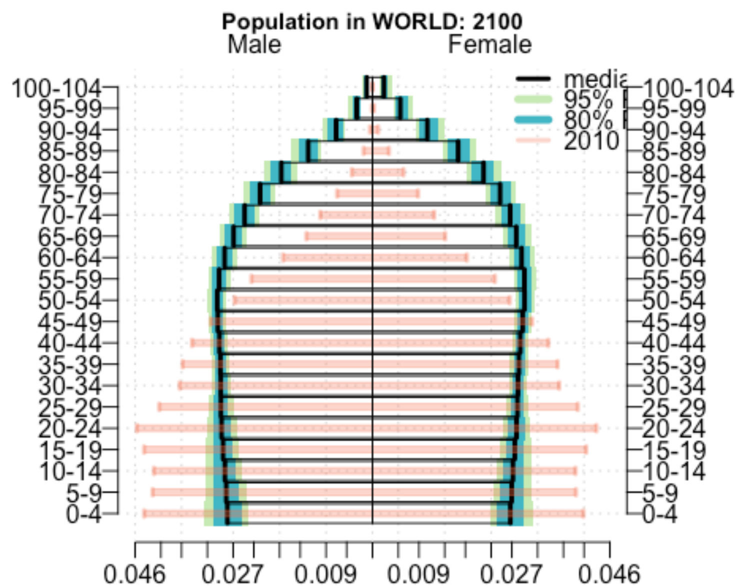


Fig. 1 Total population (1950 to 2100)

Fig. 2 Population by age and sex (2010 and 2100)

Fig. 1 shows world population peaks mid-century (around 2060) and then almost plateau's out. This is consistent with UN population projections ("World Population to", 2004). Further, Fig. 2 shows world population by age and sex for 2010 and 2100. Two important points noted here – in both years the proportion of females is roughly equal to that of males. Hence, there is no projected gender disparity in population between now and future. Secondly, the figure shows that world population is gradually ageing. There is a much higher proportion of people aged 50+ years in 2100 than it is now. This indicates that dependency ratios (and therefore elderly health-care costs) will be much higher in the future than they are now.

Regional Trends – Europe, Asia and Latin America

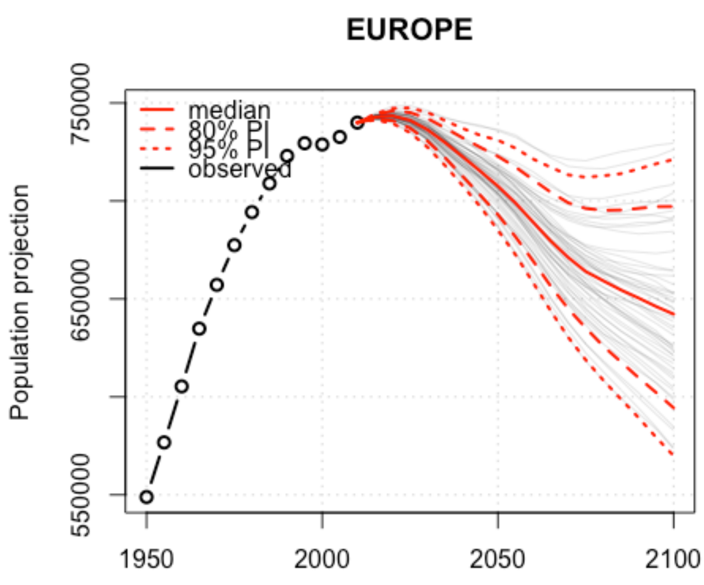


Fig. 3 Total population (1950 to 2100)

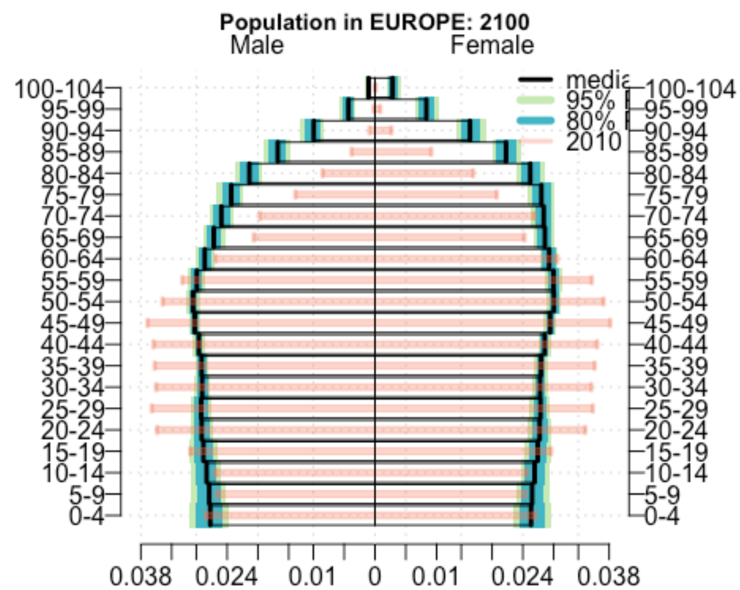


Fig. 4 Population by age and sex (2010 and 2100)

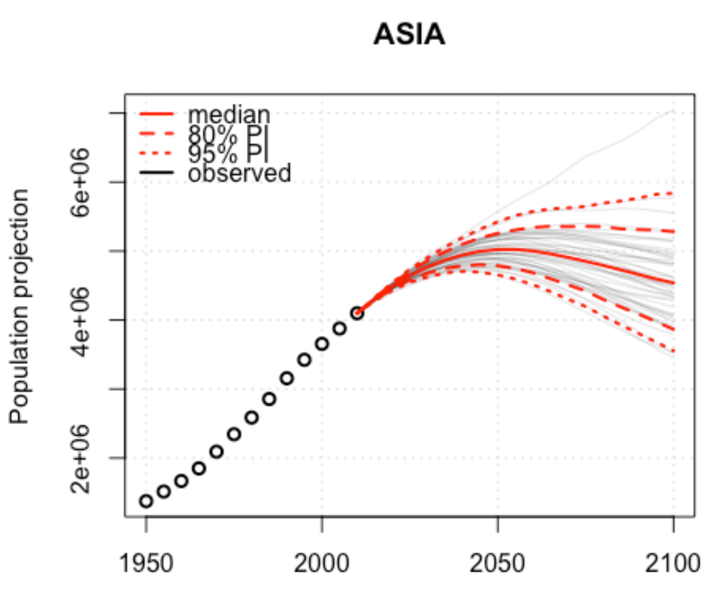


Fig. 5 Total population (1950 to 2100)

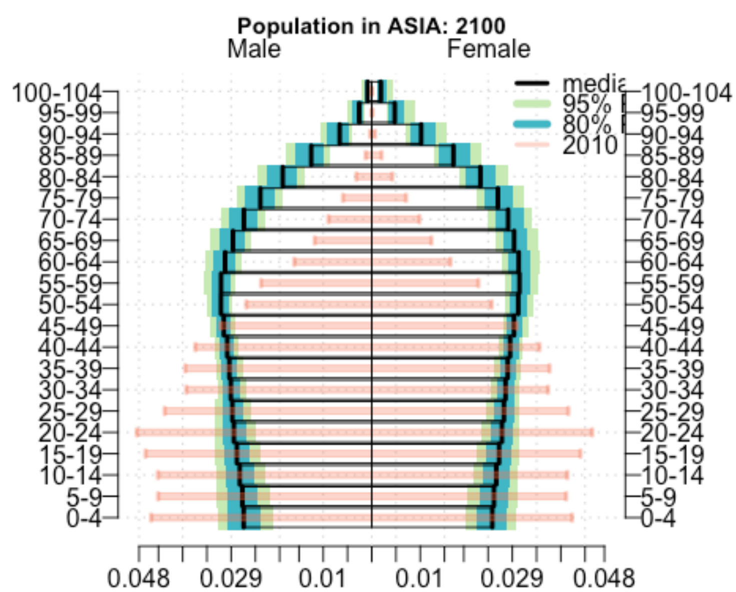


Fig. 6 Population by age and sex (2010 and 2100)

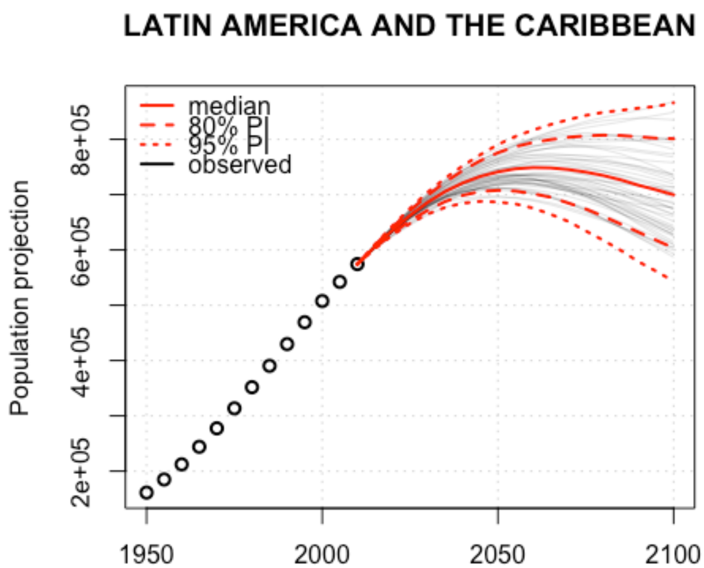


Fig. 7 Total population (1950 to 2100)

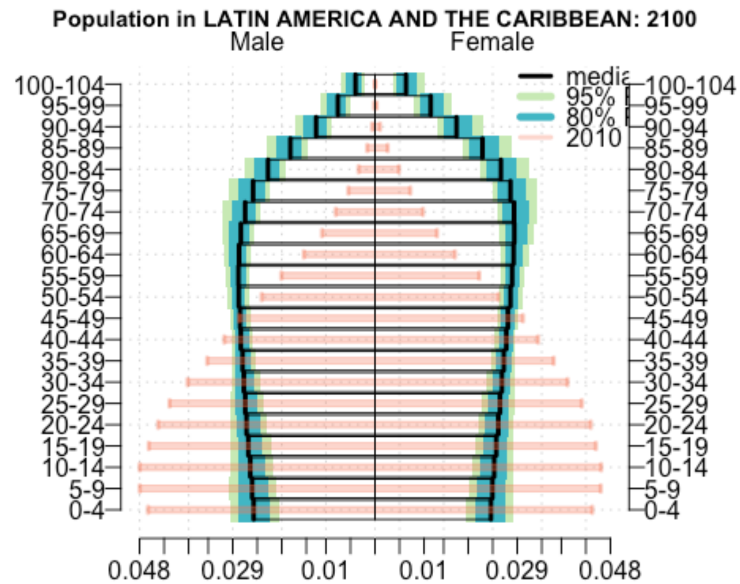


Fig 8. Population by age and sex (2010 and 2100)

The figures show a population decline in Europe, Asia and Latin America towards the end of this century. Population decline is most obvious for Europe, hence, this will play a major role in immigration policies for most European countries in the future. Also as expected, an ageing population in noticed in all three regions.

Regional Trends – Africa and North America

On the other hand, figures (please check overleaf) show a dramatic increase in population for Africa and North America by 2100. These projections not only impact the immigration policies for countries in Africa and North America, but they also highlight the importance of strong fertility-related and infrastructure-related policies in the coming decades for population control. Ageing population is expected in both regions by the end of this century.

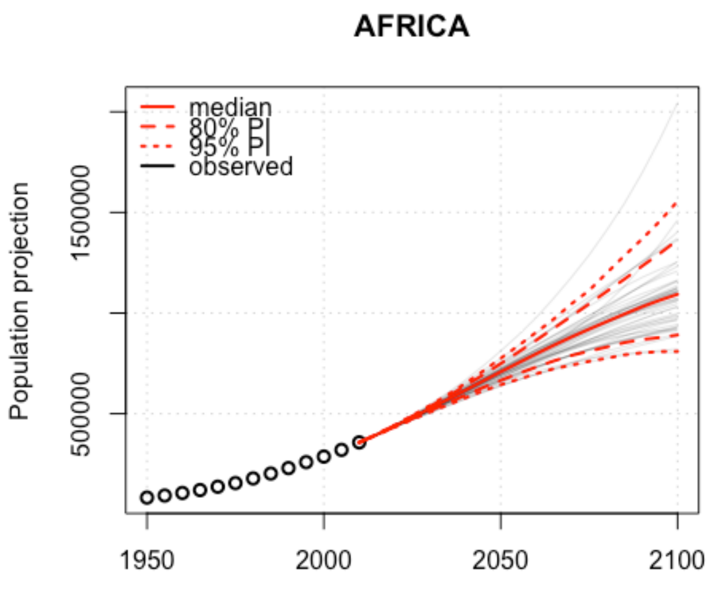


Fig. 9 Total population (1950 to 2100)

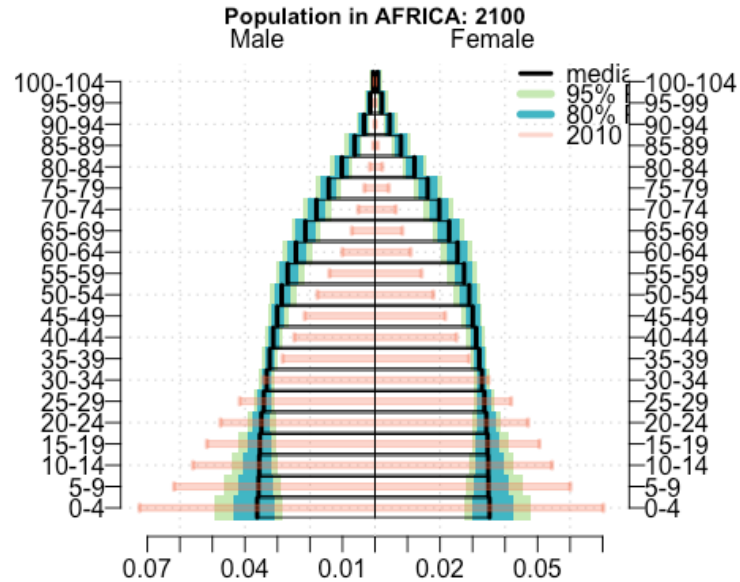


Fig 10. Population by age and sex (2010 and 2100)

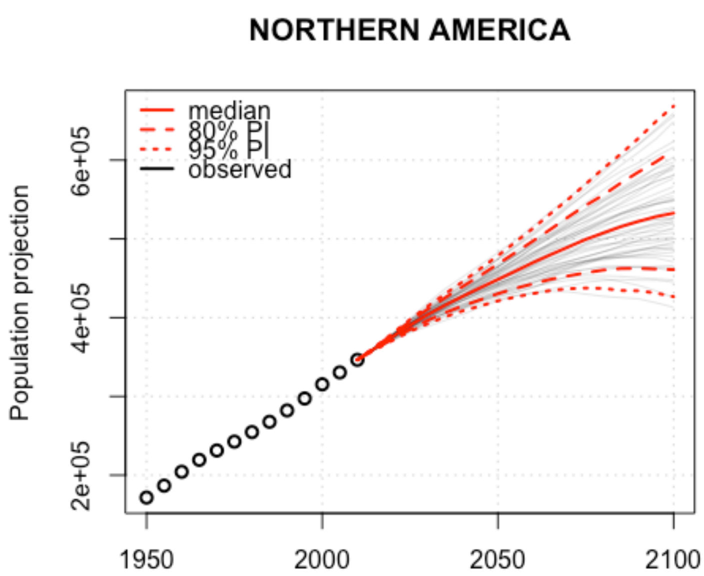


Fig. 11 Total population (1950 to 2100)

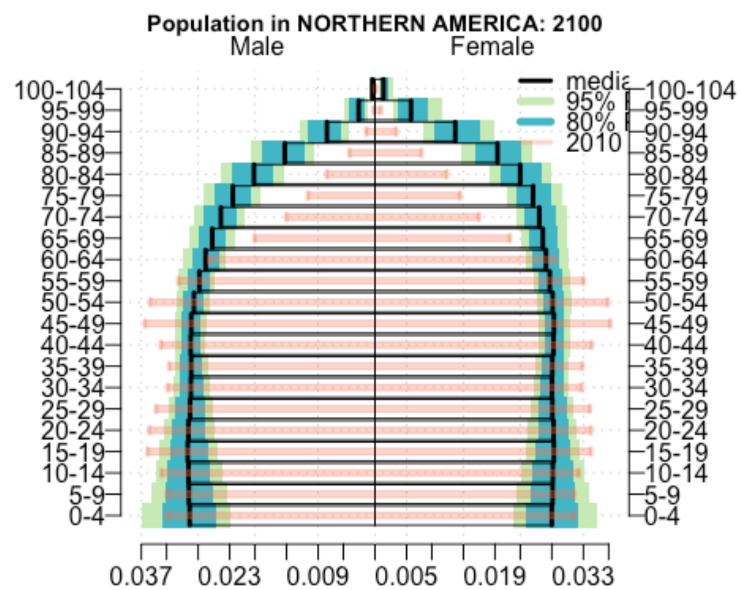
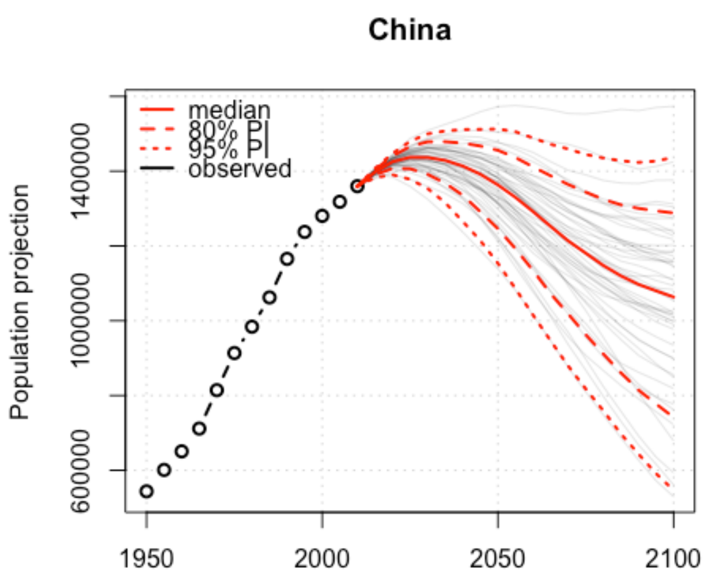


Fig 12. Population by age and sex (2010 and 2100)

Country-Specific Trends — China and India

For further analysis, we also retrieved population projections information at country level. The following figures are projections for China and India. As expected, given the decrease in population of Asia, the population of China is projected to decrease. For India, the population is expected to increase mid-century followed by a slight decline. From the figures, ageing population is expected in both countries by 2100.



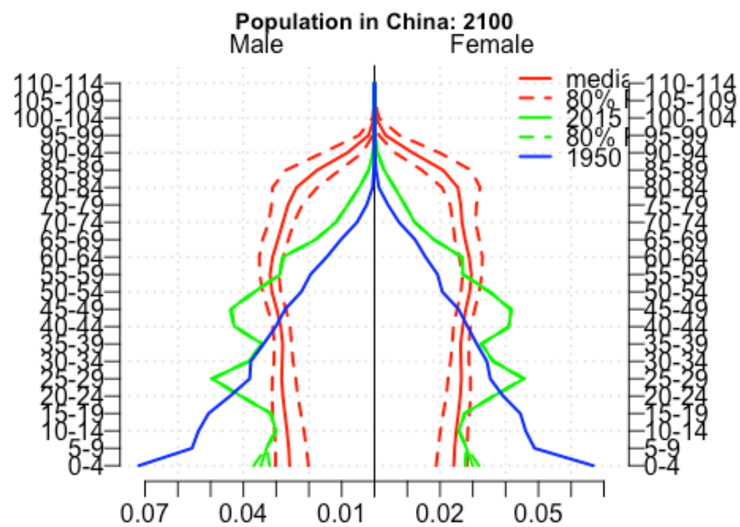
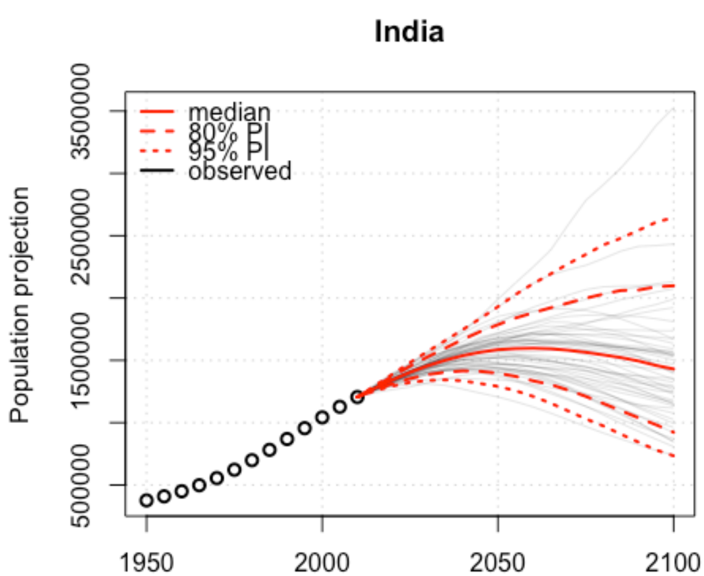


Fig. 13 Total population (1950 to 2100)

Fig. 14 Pop. by age and sex (1950, 2015, 2100)



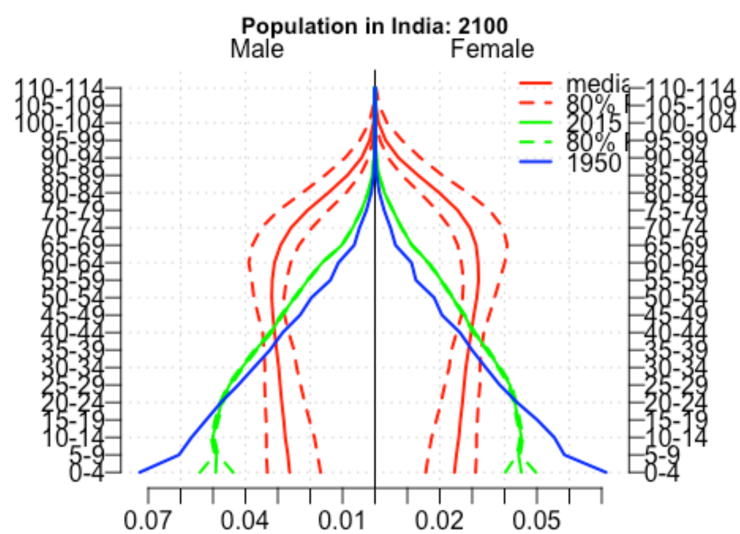


Fig. 15 Total population (1950 to 2100)

Fig. 16 Pop. by age and sex (1950, 2015, 2100)

Conclusion

For our Capstone assignment, we estimated aggregate, age-specific, and sex-specific population projections until 2100 for 200+ countries. Although our results section highlight only a few region-specific and country-specific examples, population projections for 200+ countries can be obtained using the code we have provided with this assignment. Three major trends and their policy implications are as follows:

- **A globally ageing population**

  Ageing population is consistently evident is all our results, which will increase dependancy ratios in the future. An ageing population will negatively impact healthcare and retirement planning systems, hence, it is recommended that all government actively seek out solutions to this not-too-distant issue.

- **Massive decrease in population in Europe**

  Europe is projected to undergo massive population decrease by end of the century. Coupled with an ageing population, this is have a considerable impact on European labour force. Employment-based immigration can be possible solution for this situation.

- **Massive increase in population in Africa**

  Due to high fertility rates, Africa is projected to have a high increase in population by the end of the century. Coupled with weak infrastructure and low unemployment, this will negatively impact African economies if correct policy measures (family planning) are not undertaken.

For further research, we suggest using hierarchical Bayesian estimates for estimating migration-related parameters as well for more accurate population projections.

References

(2016). Uses and Limitations of Population Projections, **National Records of Scotland.**

Retrieved from http://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/population/population-projections/uses-and-limitations-of-population-projections

Kaneda, Toshiko, & Bremner, Jason. (2014). Understanding Population Projections: Assumptions Behind Numbers. **Population Reference Bureau.** Retrieved from http://www.prb.org/pdf14/understanding-population-projections.pdf

(2000). Life expectancy at birth. Charting the Progress of Populations. United Nations. Retrieved from http://www.un.org/esa/population/publications/charting/9.pdf

Menozzi, Clare. (2015). International Migration Report 2015. United Nations Department of Economics and Social Affairs. Retrieved from http://www.un.org/en/development/desa/population/migration/publications/migrationreport/docs/MigrationReport2015_Highlights.pdf

(2004). World Population to 2300. United Nations Department of Economics and Social Affairs. Retrieved from http://www.un.org/esa/population/publications/longrange2/WorldPop2300final.pdf

Sevcikova, Hana, Raftery, Adrian E., & Gerland, Patrick. (2013). Bayesian Probabilistic Projections: Do it Yourself. **United Nations Statistical Commission and Economic Commission for Europe.**