

Case Study – Life Expectancy in Barcelona

Data Analysis and Visualization in R (IN2339)

Katharina Brenner, Julius Kasiske, Niklas Bühler, Lara-Selin Avcı

2022-01-23

Motivation

Policy makers aim to maximize prosperity among their constituents, after all, caring for voters also maximizes their chances for reelection. The data we are given for the city of Barcelona gives us a unique chance to closely examine one metric that we believe best serves as a proxy for maximizing prosperity of constituents: life expectancy! Our case study therefore focuses on some possibilities of increasing life expectancy for the citizens of Barcelona by elaborating on four claims. At the core will be the evaluation of the effect unemployment has on life expectancy. Analyzing unemployment and life expectancy and their correlation will allow policy makers to create actionable frameworks for managing unemployment respectively. We will directly look at one of such frameworks for unemployment, social mobility, in order to hopefully lay out a path for politicians to maximize prosperity among their voters in Barcelona.

Data Preparation

In this section, we perform the needed data preparation steps needed for the analysis.

Population Data

```
# Reduce unemployment table to Year == 2014
unemployment <- unemployment0[Year == 2014 & `Neighborhood Name` != "No consta"]

# Bring unemployment table in neat format
unemployment <- unemployment[, sum(Number)/12, by = .(`Neighborhood Name`, Gender)] %>%
  dcast(., ... ~ Gender)
colnames(unemployment) <- c("neighborhood", "F", "M")
unemployment <- melt(unemployment, id.vars = "neighborhood", variable.name = "gender",
  value.name = "unemployment")

# Reduce life expectancy table to years 2010-2014
lifeExpectancy <- lifeExpectancy0[, .(Neighborhood, `2010-2014`, Gender)] %>%
  dcast(., Neighborhood ~ Gender, value.var = "2010-2014")

# Bring life expectancy table in neat format
colnames(lifeExpectancy) <- c("neighborhood", "F", "M")
lifeExpectancy <- melt(lifeExpectancy, id.vars = "neighborhood", variable.name = "gender",
  value.name = "lifeExpectancy")

# Reduce population table to Year == 2014
population <- population0[Year == 2014]

# Bring population table in neat format
population <- population[, sum(Number), by = .(Neighborhood.Name, Gender)] %>%
  dcast(., Neighborhood.Name ~ Gender, value.var = "V1")
```

```
colnames(population) <- c("neighborhood", "F", "M")
population <- melt(population, id.vars = "neighborhood", variable.name = "gender",
  value.name = "population")
```

```
# Merge all tidy tables into one allTables table
allTables <- merge(unemployment, population, by = c("neighborhood", "gender"))
allTables <- merge(allTables, lifeExpectancy, by = c("neighborhood", "gender"))
```

```
# Calculate unemployment rate in merged table
allTables <- allTables[, unemploymentRate := unemployment/population]
```

```
# Remove NA's
allTables <- na.exclude(allTables)
```

We end up with the following population related data:

```
rbind(head(allTables, 1), tail(allTables, 1))
```

```
##      neighborhood gender unemployment population lifeExpectancy
## 1: Baró de Viver      F      93.33333      1281           85.7
## 2: les Tres Torres    M     212.16667      7507           83.9
##      unemploymentRate
## 1:      0.07285974
## 2:      0.02826251
```

Transportation Data

```
# Obtain stops of relevant means of transportation
transportation <- transportation0[Transport %in% c("Underground",
  "Railway (FGC)", "Tram", "RENFE")]
busstops <- busstops0[Transport == "Day bus stop"]
```

```
# Merge data from all means of transportation into one table
busstops <- busstops[, data.table(
  table(Neighborhood.Name, Transport))][Neighborhood.Name != ""]
transportation <- transportation[, data.table(
  table(Neighborhood.Name, Transport))][Neighborhood.Name != ""]
prepList <- list(bustops, transportation)
transportation <- rbindlist(prepList)
```

```
# Bring transportation table in neat format
colnames(transportation) <- c("neighborhood", "transport", "stations")
transportation_casted <-
  dcast(transportation, neighborhood ~ transport, value.var = "stations")
transportation_casted[is.na(transportation_casted)] = 0
```

```
# Merge with relevant parts of the population data
allTablesCopy <- copy(allTables)
allTablesTransport <- allTablesCopy[, c("lifeExpectancy", "unemploymentRate")
  := .(NULL, NULL)]
allTablesTransport <- allTablesTransport[, .(unemploymentAll = sum(unemployment),
  populationAll = sum(population)),
  by = neighborhood]
allTablesTransport <- merge(allTablesTransport, transportation_casted,
  by = "neighborhood", all.x = T, all.y = F)
```

```

# Calculate total station counts
allTablesTransport[, Sum := `Day bus stop` + RENFE + Underground + Tram + `Railway (FGC)`]

# Calculate rates for unemployment and transportation stations
allTablesTransport02 <-
  allTablesTransport[, c("unemploymentRate", "transportDensity")
    := .(unemploymentAll/populationAll, Sum/populationAll)]

# Rename columns
colnames(allTablesTransport02) <-
  c("neighborhood", "unemployment", "population", "bus", "renfe", "railway(fgc)",
    "tram", "subway", "total", "unemploymentRate", "totalDensity")

# Calculate rates per means of transport
allTablesTransport02[, c("busDensity", "renfeDensity", "railwayDensity",
  "tramDensity", "subwayDensity")
  := .(bus/population, renfe/population, `railway(fgc)`/population,
    tram/population, subway/population)]

# Melt the transportation data table
allTablesTransportMelted <-
  melt(allTablesTransport02[, .(neighborhood, busDensity, renfeDensity, railwayDensity,
    tramDensity, subwayDensity, totalDensity,
    unemploymentRate)],
    id.var = c("neighborhood", "unemploymentRate"),
    variable.name = "transport", value.name = "transportDensity")

```

We end up with the following transportation related data:

```

rbind(head(allTablesTransportMelted, 1), tail(allTablesTransportMelted, 1))

```

```

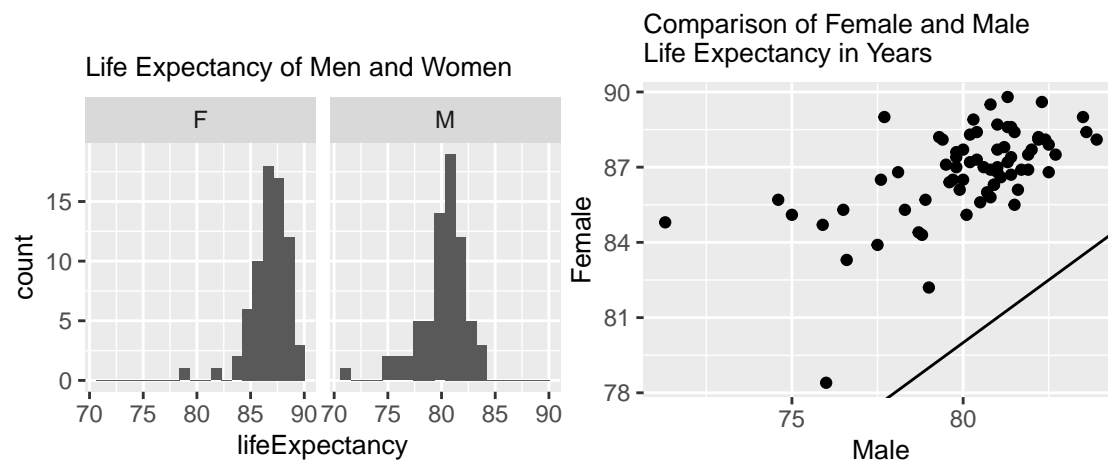
##      neighborhood unemploymentRate   transport transportDensity
## 1:  Baró de Viver      0.07067043 busDensity      0.002418380
## 2:  les Tres Torres    0.02873688 totalDensity    0.001856321

```

Data Analysis

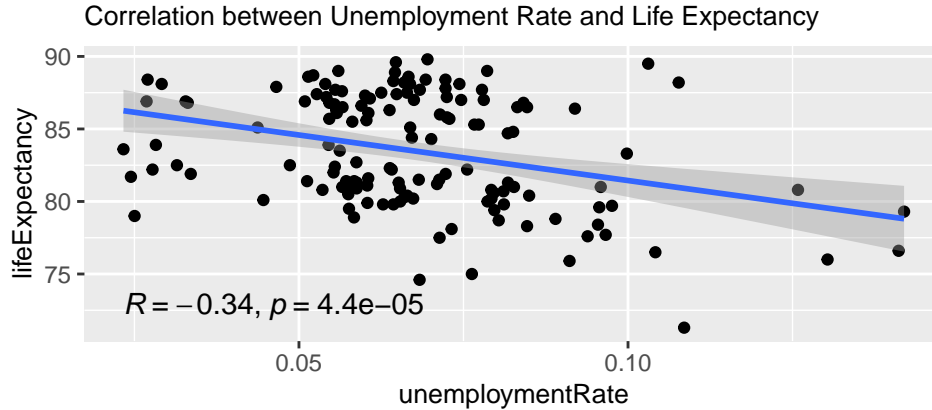
1. On Average, Women Live Longer than Men

Evaluating life expectancy requires understanding how the metric is distributed within the city. As is shown below, women in Barcelona live longer than men do. As such, Barcelona presents no exception to the general rule observed around the globe.



2. Higher Unemployment Rates Might Lead to Lower Life Expectancies

Establishing an association between unemployment and life expectancy is key to our study. Plotting life expectancy against unemployment indicates that our primal hypothesis holds: a correlation exists, enabling us to further reason about a possible causation. Performing a pearson test on the correlation further indicates statistical significance of our finding.



3. The Correlation Between Unemployment Rates and Life Expectancies is Strong for Men and Almost Non-existent for Women

However, stratifying the above correlation by gender forces us to adapt our primal hypothesis, as gender seems to be a confounding factor to the relationship studied above. This is shown in the plot below, where the correlation between unemployment rate and life expectancy is visualized by gender.

It can be seen that the effect is almost non-existent for women, and the negative correlation is even stronger for men than for the whole population in general.



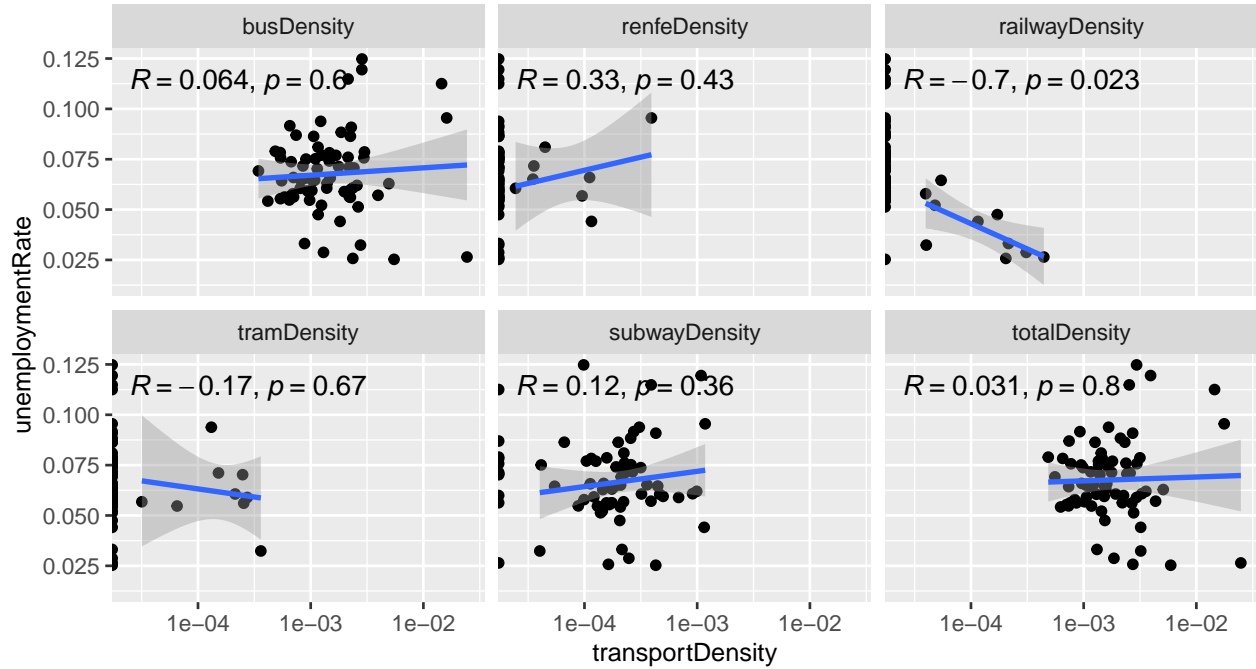
These findings constitute only correlations. However, the association also makes causal sense. Employment grants access to health care and provides the means for stable shelter and healthier, more nutritious eating habits, thus minimizing the risks from undetected diseases, accidents, starvation, obesity and much more. Thus, minimizing unemployment will go a long way in maximizing life expectancy among populations.

4. With the Possible Exception of Railway, Public Transport Density Does Not Have an Impact on Unemployment Rate

Minimizing unemployment is a lot more actionable than optimizing life expectancy. One approach to managing unemployment is access to mobility - the more connected people are to other areas, the larger their addressable job market. We thus want to analyze how the number of stations per capita (as a proxy for access to mobility) influences unemployment.

When stratifying by means of transport, it becomes visible that the density of railway stations in a neighborhood does indeed correlate quite strongly with unemployment rate. Thus, having more access to railway stations might increase social mobility in the form of higher employment rates in a given neighborhood.

Correlations Between Public Transport Density and Unemployment Rate, Split by Means of Transport



Conclusion

We hypothesized that providing access to mobility would decrease unemployment and indirectly affect prosperity in Barcelona. However, stratifying by means of public transport shows that that only holds true for railway access.

Thereby, in order to combat structural inequality in Barcelona, decision makers could strategically build more railway stations in neighborhoods with high unemployment rates. This could lead to an increase in social mobility in the form of employment and possibly also increase the male life expectancy in these neighborhoods.