Module 3

Day 2: Maps!!!

Recap of Last Week

- Fixed Effects
- Differences-in-Differences

Recap: Fixed Effects

- Fixed effects are a tool that allow us to control for unobserved variation through the use of dummy variables
- A **Factor Model** is an econometric model that uses fixed effects to control for the variation for the variables that are not of interest
- Pros
  - Can correct for omitted variable bias
  - Low computational cost
- Cons
  - Need large amounts of data for your variables of interest to keep statistical significance
  - Might lead to over fitting

Recap: Fixed Effects pt. 2

Suppose we have annual wage data for 20,000 individuals that spans from 2000-2017. We want to investigate what effect does education have on wages, so we estimate the following model

$$WAGES_{it} = \beta_0 + \beta_1 EDUC_{it} + \epsilon_{it}$$

Where $WAGES_{it}$ is the annual wages and $EDUC_{it}$ is the years of education individual i has in time t

- Is this Panel, Cross Sectional, or Time Series data?
- What are some ommited variables?

Recap: Fixed Effects pt. 3

We can control for all of those omitted variables by using individual and fixed effects!

$$WAGES_{it} = \beta_0 + \beta_1 EDUC_{it} + \delta_i + \lambda_t + \epsilon_{it}$$

Where δ_i is a dummy variable for each individual!

- How many total dummy variable are there?
- What happens if we include time fixed effects (dummy variables for each year)?

Recap: Differences-in-Differences

- The ingredients
  - Treatment - policy intervention
  - Control group
  - Treated group
- Look at the difference between the marginal effect between the treated and control group to back out the impact of the treatment

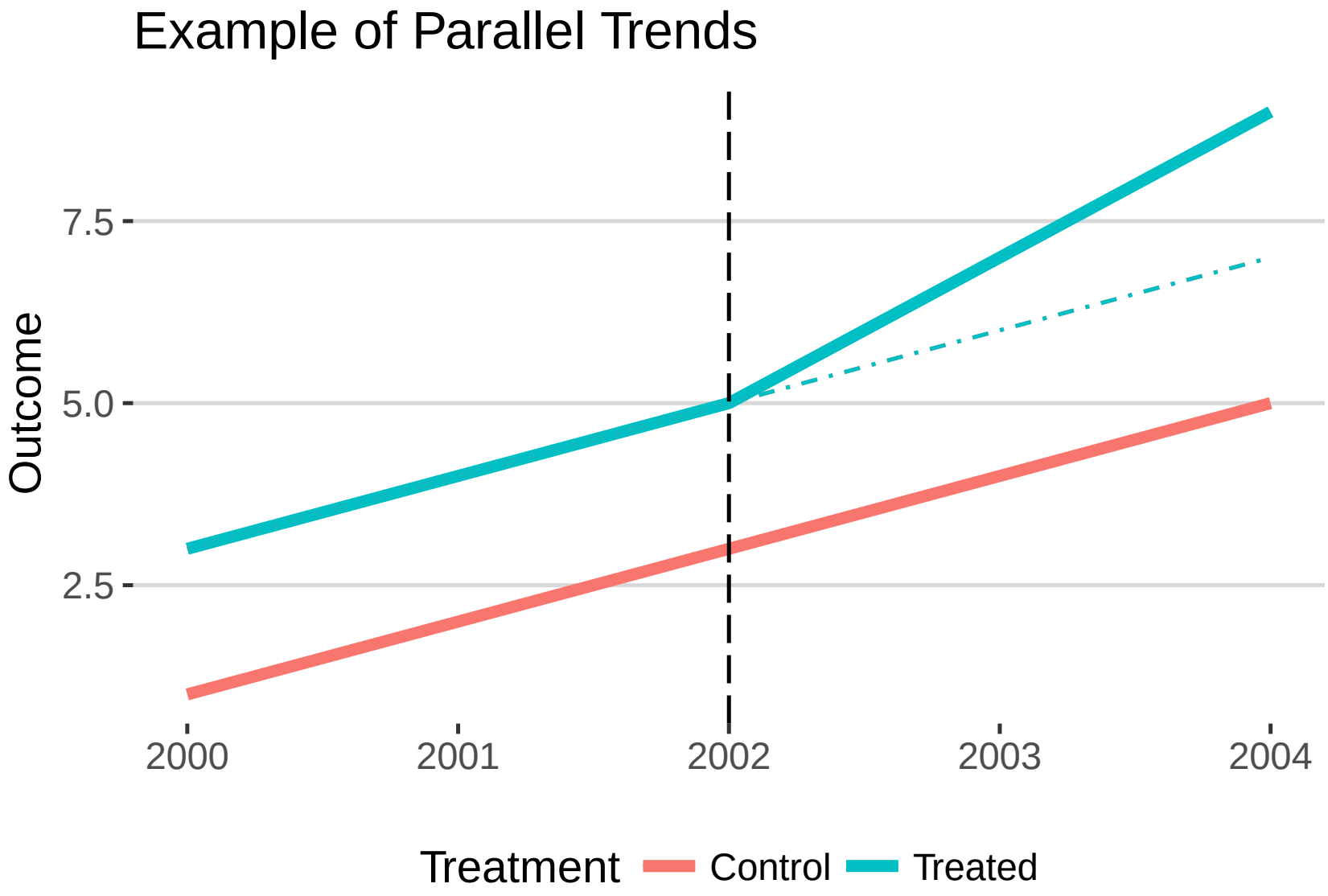
Recap: Differences-in-Differences pt. 2

- Pros
  - Estiamtes the causal effect of the treatment in a rigirous and well understood manner
  - Easy and fast to estimate
- Cons
  - Very specific type of data
  - Evidence of parallel trends
  - Only describes the effect of treatment not how the mechanism works

Recap: Differences-in-Differences pt.3

- Last time we recreated Card and Kruger (AER 1994)
- The authors investiagted if an increase in the mimum wage leads to a decrease in employment
- The ingrediants
  - Treatment - law that increased minimum wages
  - Control group - New Jersey
  - Treated group - Pennsylvania
- They find that there was no effect of the minumum wage increase on employment!
- Any thoughts? What did they do well? What could have been improved?

Recap: Differences-in-Differences pt.4



Game Plan for the Day

- Investigate some development data
  - Measuring growth vs. measuring development
  - Precanned objects in functions
  - Relationship between wealth and development
- Plotting maps in R
  - Create maps with ggplot
  - Incoporate data with maps

Packages with Data Frames in Them

- Some packages include sample data frames to play around with when you install it
- Examples include Starwars Movie characters, stock prices, and movie ratings
- Precanned data frames allow you skip loading data in!

Loading in Data

```
library(gapminder)
world_data <- gapminder
head(world_data)
```

```
## # A tibble: 6 x 6
##   country     continent  year lifeExp      pop gdpPercap
##   <fct>       <fct>     <int>   <dbl>    <int>     <dbl>
## 1 Afghanistan Asia       1952    28.8  8425333       779
## 2 Afghanistan Asia       1957    30.3  9240934       821
## 3 Afghanistan Asia       1962    32.0 10267083       853
## 4 Afghanistan Asia       1967    34.0 11537966       836
## 5 Afghanistan Asia       1972    36.1 13079460       740
## 6 Afghanistan Asia       1977    38.4 14880372       786
```

Quick Glance at the data

- What frequency is data reported at?
- What variable is measuring quality of life (i.e Development)?

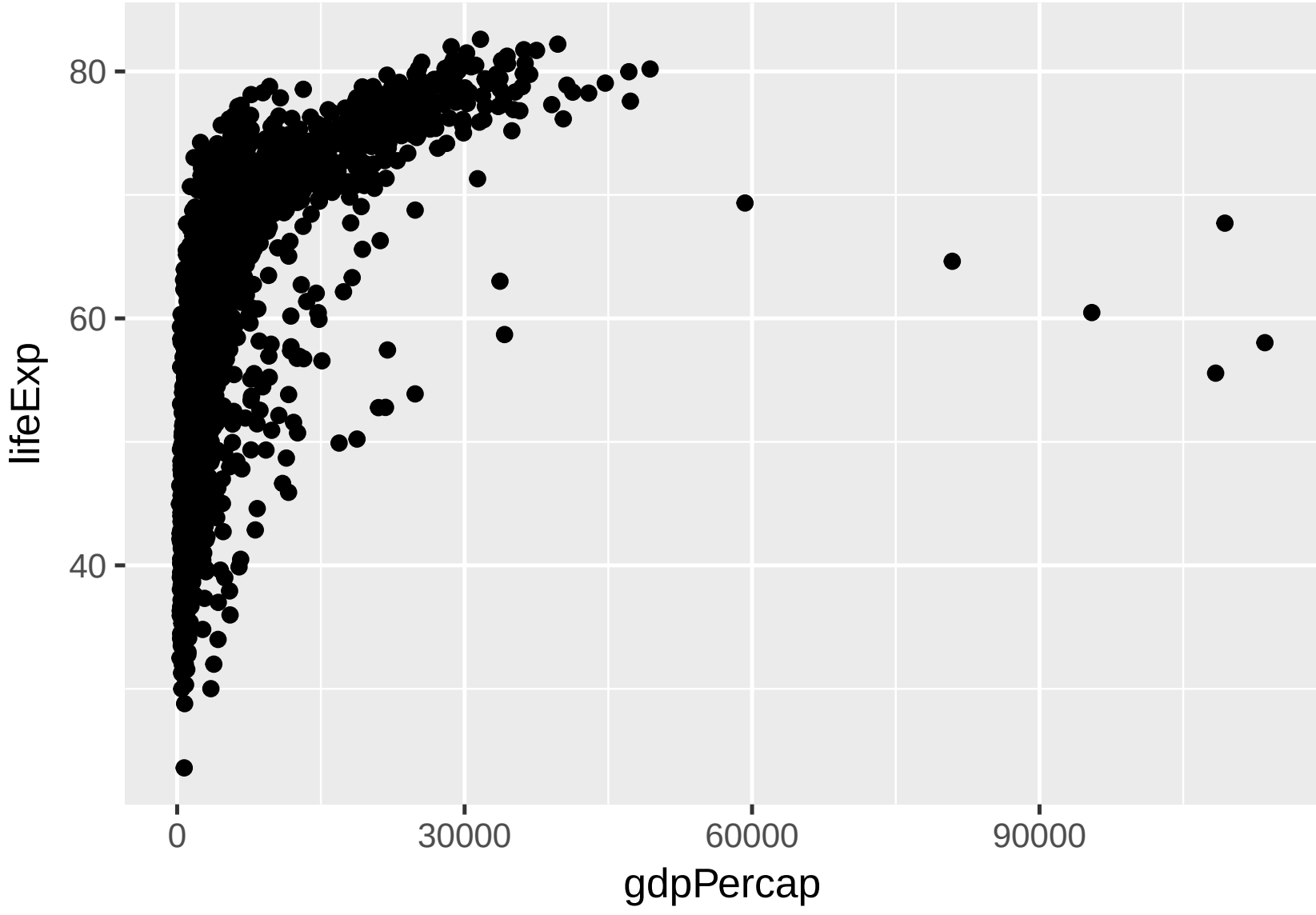
Development vs. Growth

- **Economic Growth** is the study of how economies grow through technology, research, and investments innovations.
- What recent technological innovations do you think that have attributed to growth in the American economy?
- **Development Economics** is the study of how nations improve the economic, political, and social well-being of its people.
- How can we measure development?

The Relationship between Economic Growth and Development

- Is there a correlation between a country's wealth and how "developed" it is?
- How much of a country's development is due to the country's wealth?
- This is our research question for the day!
- How would we investigate this relationship with our data?

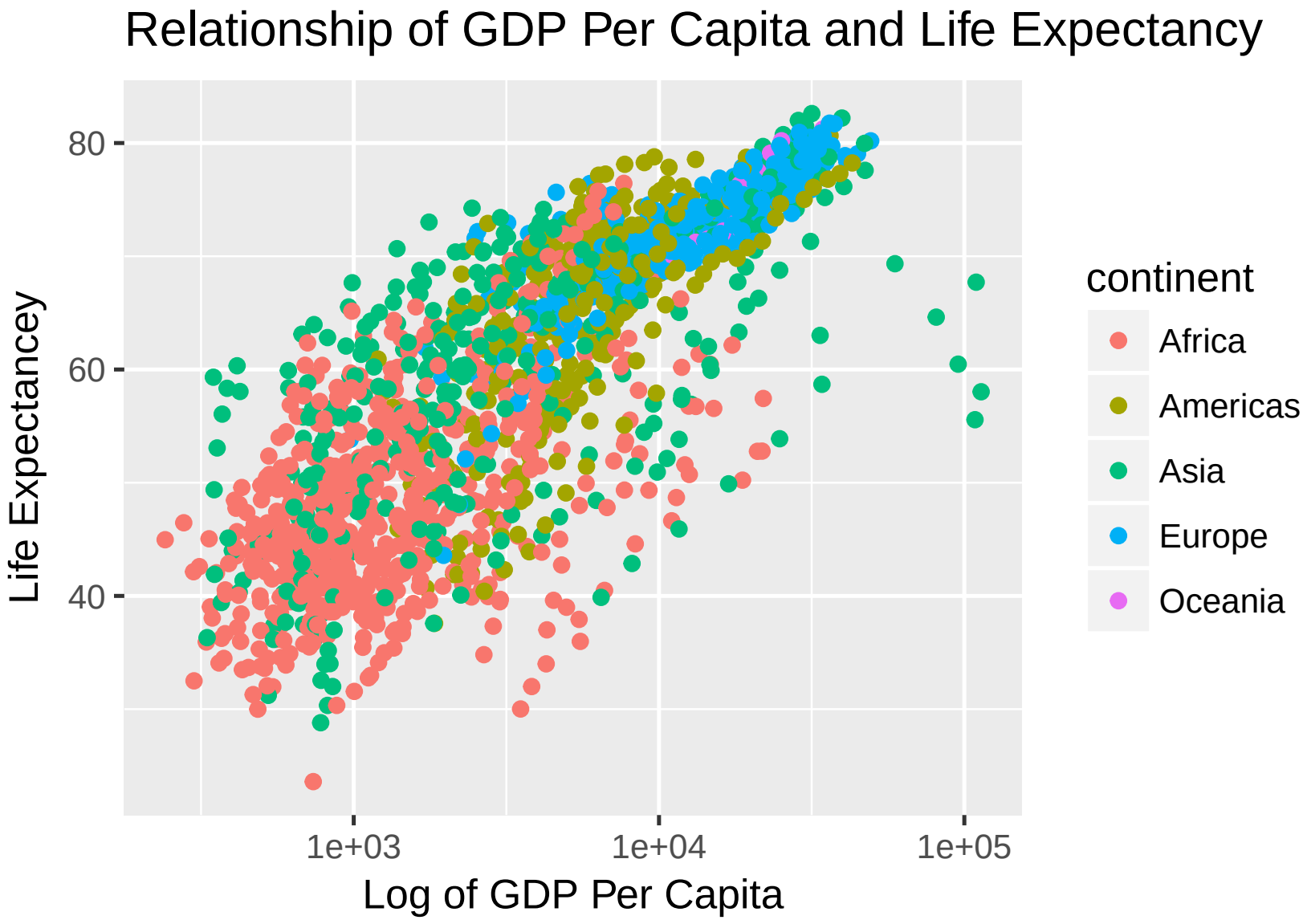
Plotting the relationship between GDP Per Capita and Life Expectancy



In Class Exercise 1: Make a pretty plot!!!

- Y'all are now plotting pros!
- Fix up the previous plot and differentiate the observations by contitinet.
- I would suggest to use scale_x_log10(), if you do not know what this function does please google it or type ?scale_x_log10() into you counsle

In Class Exercise 1: Solution



Creating a Benchmark Model

```r
benchmark <- lm(lifeExp ~ gdpPercap, data = world_data)
summary(benchmark)
```

```
##
## Call:
## lm(formula = lifeExp ~ gdpPercap, data = world_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -82.754  -7.758   2.176   8.225  18.426
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 5.396e+01  3.150e-01  171.29   <2e-16 ***
## gdpPercap   7.649e-04  2.579e-05   29.66   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
##
```

Interpreting and Performance our Benchmark Model

- What dose a $1000 increase in GDP Per Capita have on life expectancy?
- How "good" is our model? What tools did you just use?
- What about missing variables that could explain the variation in Life Expectancey?
- Our data has its limits, but econometric tool can we use to imporve the estimates?

Incorporating Time Fixed Effects

```
time_fe_reg <- lm(lifeExp ~ gdpPercap + factor(year),
                  data = world_data)
summary(time_fe_reg)
```

```
##
## Call:
## lm(formula = lifeExp ~ gdpPercap + factor(year), data =
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -66.880  -6.915   0.994   7.606  21.052
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       4.655e+01  8.161e-01  57.041  < 2e-16 *
## gdpPercap         6.721e-04  2.442e-05  27.521  < 2e-16 *
## factor(year)1957  2.064e+00  1.147e+00   1.799 0.072156 .
## factor(year)1962  3.879e+00  1.147e+00   3.381 0.000738 *
```

Did Time Fixed Effects Help?

- How do we know? What tools did you use to say yes or no?
- Still missing variables that explain the variation between countries
- Let's add country fixed effects!

In Class Exercise 2: Country Fixed Effects

- Run a regression named "country_fe_reg" with just fixed effects
- Run a regression named "both_fe_reg" with both time and country fixed effects
- Put benchmark, time_fe_reg, country_fe_reg, and both_fe_reg into a stargazer table
- Please exclude the coefficents reported on the fixed effects, but indicate what regression has which fixed effects.

In Class Exercise 2: Solution

```
benchmark <- lm(lifeExp ~ gdpPercap,
                data = world_data)
time_fe <- lm(lifeExp ~ gdpPercap + factor(year),
              data = world_data)
country_fe <- lm(lifeExp ~ gdpPercap + factor(country),
                 data = world_data)
both_fe <- lm(lifeExp ~ gdpPercap + factor(year) +
                factor(country), data = world_data)

stargazer(benchmark, time_fe, country_fe, both_fe,
          type = "text",
          covariate.labels = c("GDP Per Capita"),
          omit = c("factor"),
          add.lines = list(c("Time Fixed Effects?",
                             "No", "Yes", "No", "Yes"),
                           c("Country Fixed Effects?",
                             "No", "No", "Yes", "Yes"))
          )
```

In Class Exercise 2: Interpreting Results

- What is the "best" model?
- Why does the sign infront of the coefficent "flip"?
- Is the model with time and country fixed effects over-fitted?

Vizualizing Data with Maps

- Powerful tool
- Dipicts clustering
- We will be using another precanned data set from the ‘maps’ package

Loading in Map Data

```
world <- map_data("world")
head(world)
```

```
##         long      lat group order region subregion
## 1 -69.89912 12.45200     1     1  Aruba      <NA>
## 2 -69.89571 12.42300     1     2  Aruba      <NA>
## 3 -69.94219 12.43853     1     3  Aruba      <NA>
## 4 -70.00415 12.50049     1     4  Aruba      <NA>
## 5 -70.06612 12.54697     1     5  Aruba      <NA>
## 6 -70.05088 12.59707     1     6  Aruba      <NA>
```

Geographic Data in R

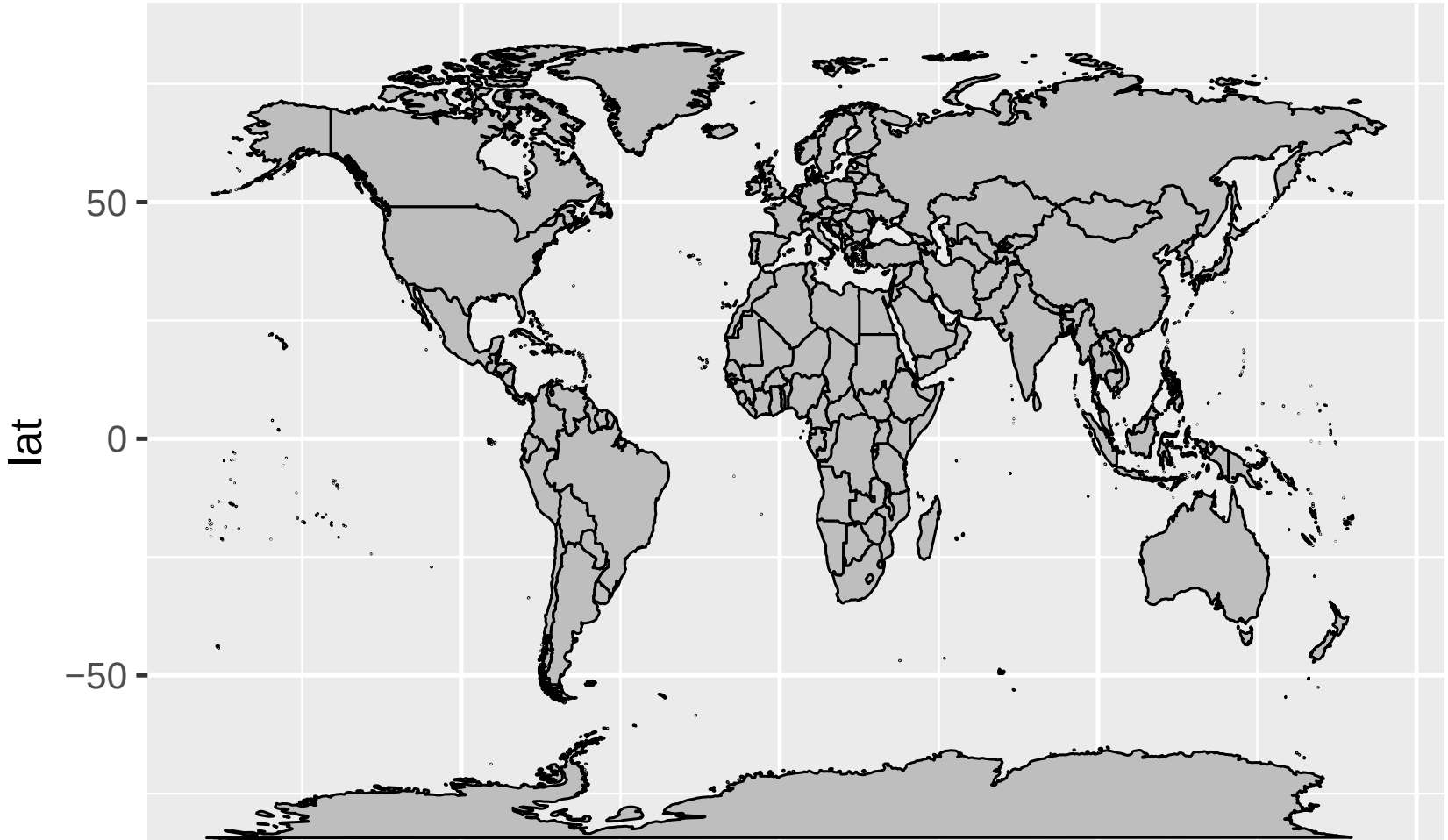
- Think of the data as a bunch of points where R is smart enough to just draw lines through the points
- Ordering matters in this type of geospatial data - so don't go too crazy on it!!!
- There are many types of way to store geographic data, and the type of data we are working with is the easiest.
- Just be careful if you are wanting to do maps in the future, most of the time you will be given shape files which are its own special thing.

Geographic Data in R pt. 2

- Mapping, in this lecture, works the exact same as a normal ggplot
- There is a new "layer" called polygon
- Note that the x variable is longitude and the y variable is lattitude
- It common for people to say "latt, long" instead of "long, latt", either way to say it is fine, but when working with geographic data 90% of the time your x variable will be long and your y variable will be latt

Our First Map

```
world %>%
    ggplot(aes(x = long, y = lat, group = group)) +
    geom_polygon(fill = "gray", color = "black",
                 size = 0.3)
```



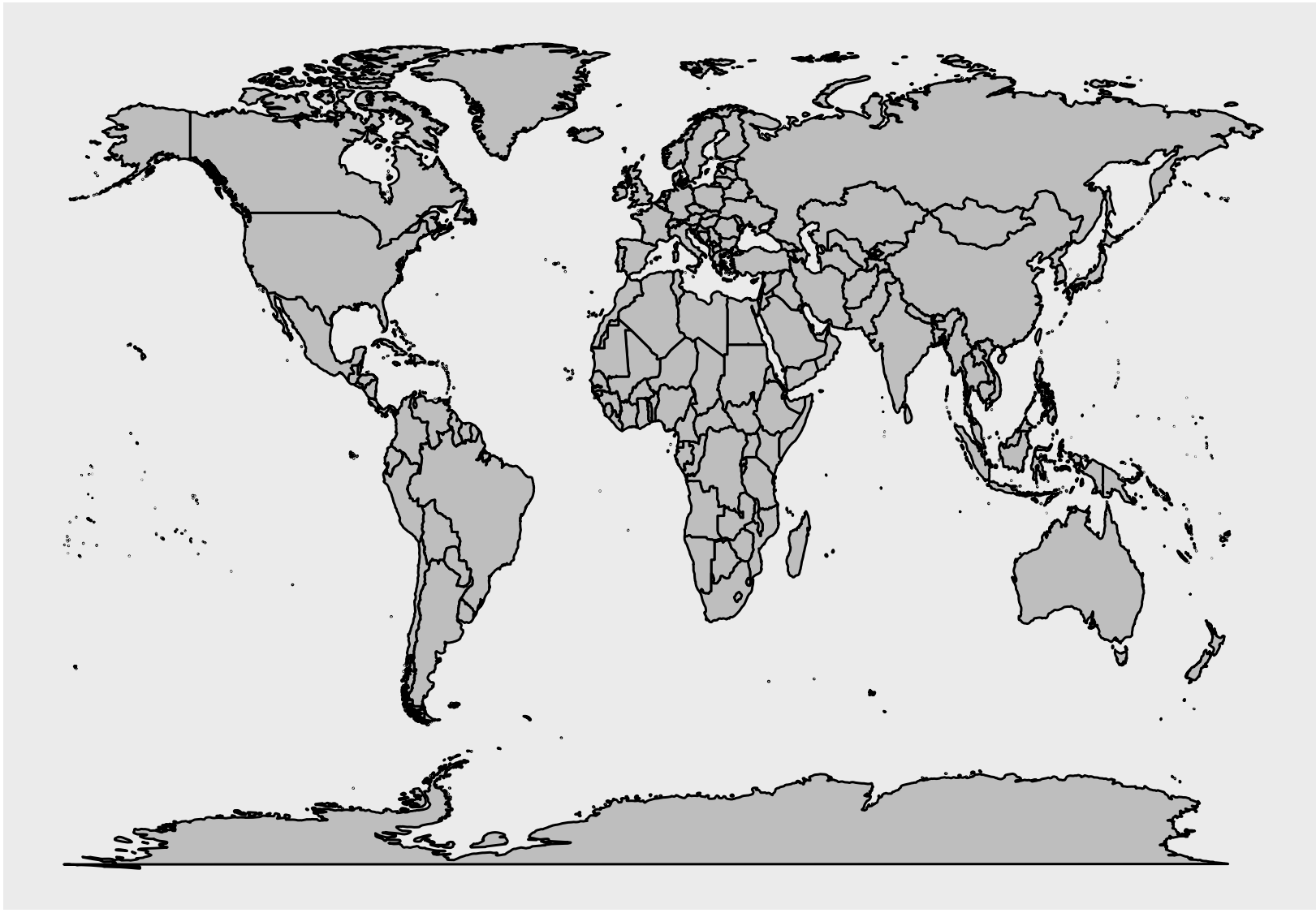
Our First Map - Improvements

- What are some improvements we can make to this map?
- Since we are using ggplots we can make the same improvements in the same manner
- The only tricky part is removing the axes

Our First Map - Removing the Axes

```
no_axes <- theme(
  axis.text = element_blank(),
  axis.line = element_blank(),
  axis.ticks = element_blank(),
  panel.border = element_blank(),
  panel.grid = element_blank(),
  axis.title = element_blank())
world %>%
    ggplot(aes(x = long, y = lat, group = group)) +
    geom_polygon(fill = "gray", color = "black",
                 size = 0.3) +
    no_axes
```

Our First Map - Pretty!



Merging Our Data with the World Map DF

- We will merge by country!
- For now let's just look at Africa

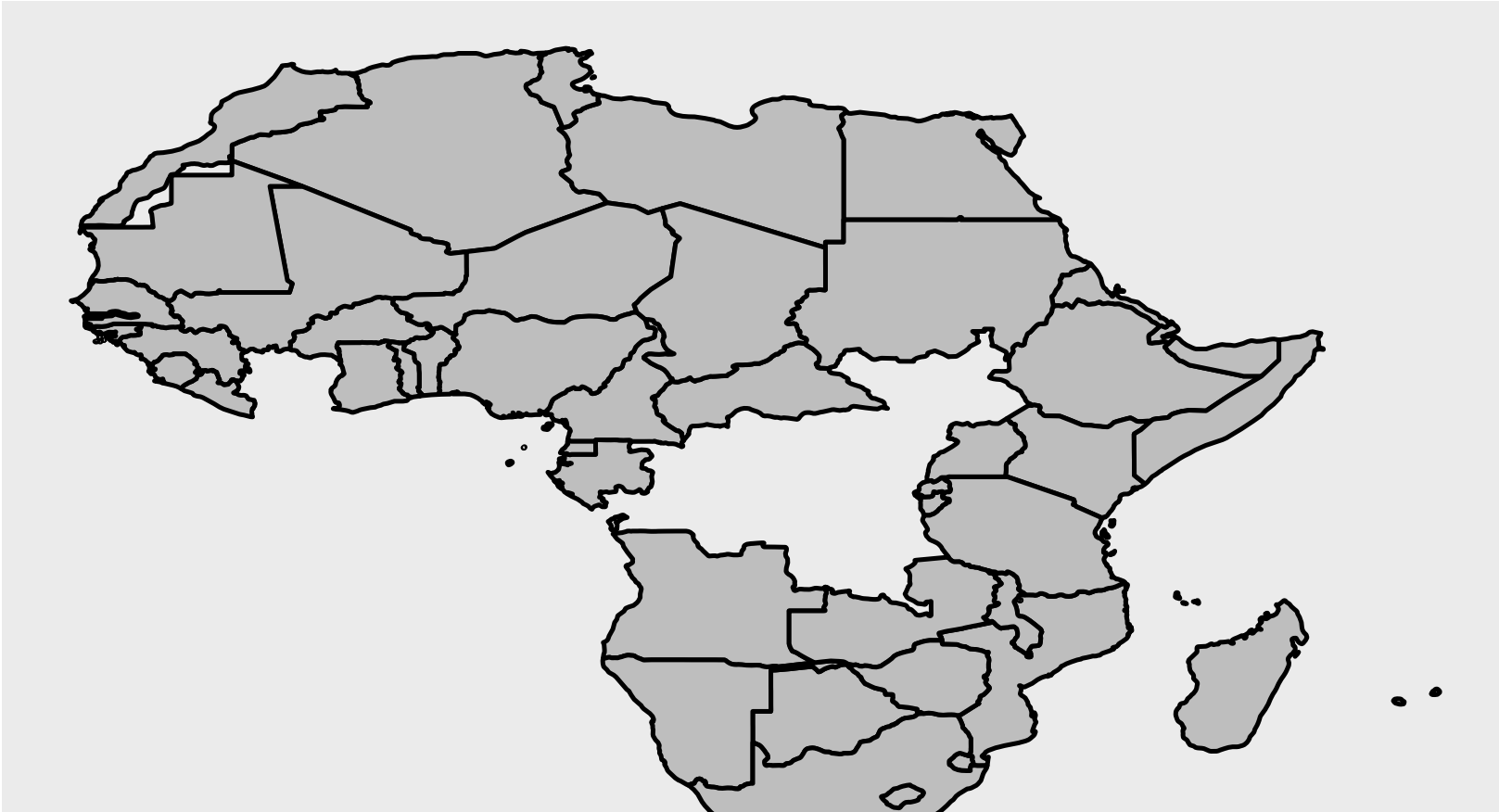
```
world_data$country <- as.character(world_data$country)

africa <- world_data %>%
  filter(continent == "Africa") %>%
  inner_join(world, by = c("country" = "region"))
```

```
## Warning: package 'bindrcpp' was built under R version 3.
```

Plotting our Merged Data

```
africa %>%
  filter(year == 2007) %>%
  ggplot() +
    geom_polygon(aes(long, lat, group = group), fill = "gre
    no_axes
```



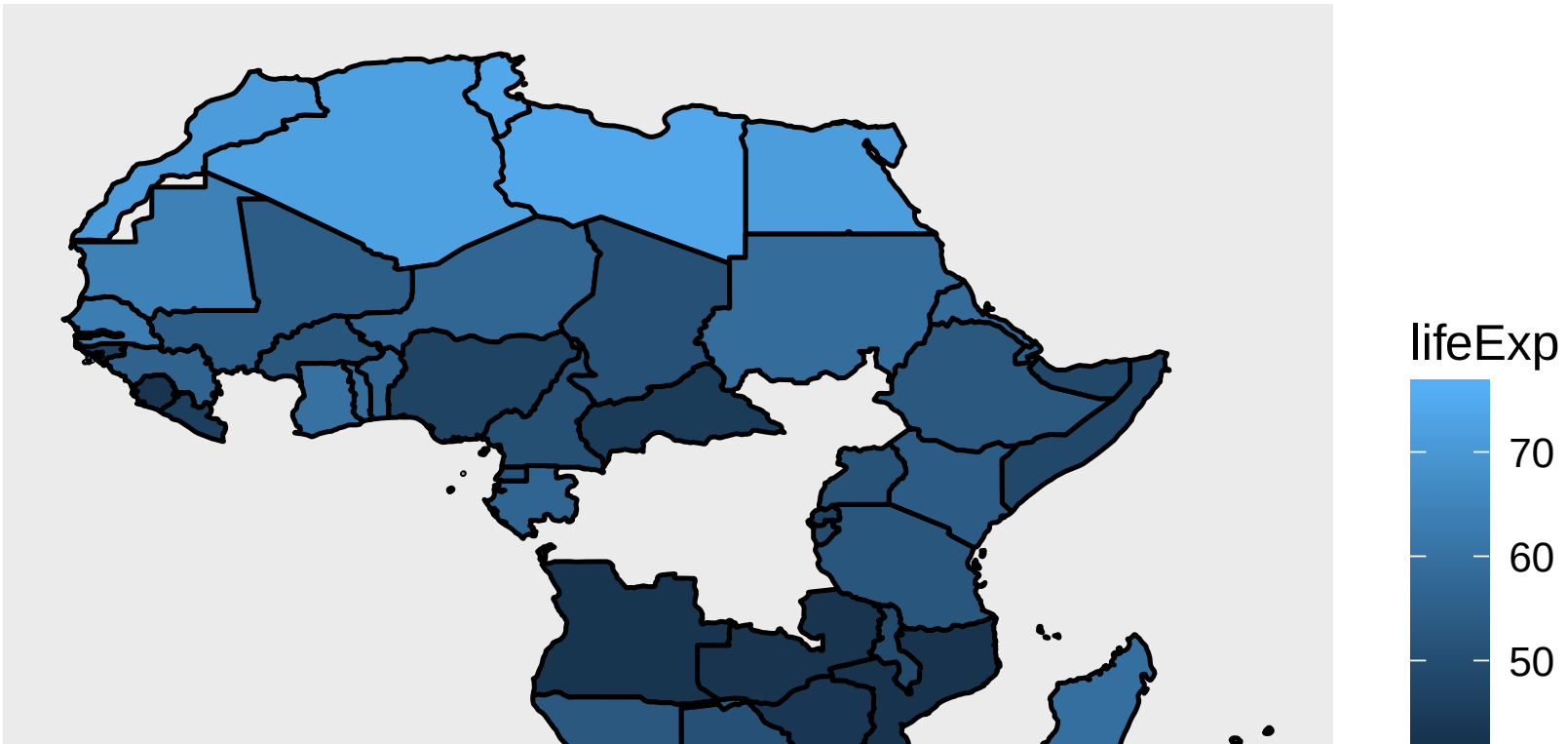
Why are there missing countries?

- Any ideas?
- Did some get dropped in the merge?
- Our data on GDP Per Capita and Life Expectancy goes back to 1954...
- This is fine for now!

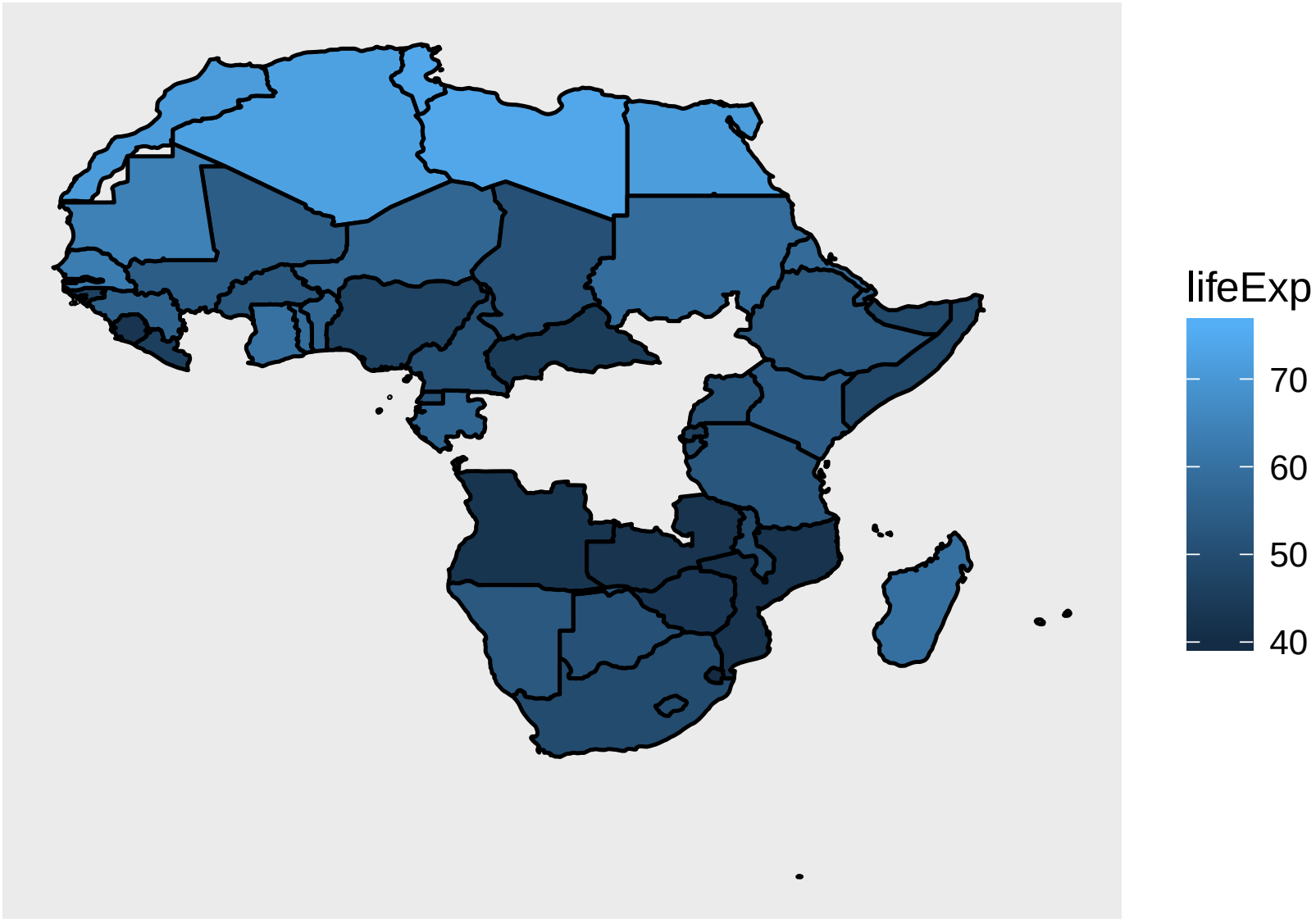
Let's add some data!

- Now let's create a heat map of life expectancy

```
africa %>%
  filter(year == 2007) %>%
  ggplot() +
    geom_polygon(aes(long, lat, group = group, fill = lifeE
    no_axes
```



Heat map of Life Expectancy



Improvements?

- How can we make the map better?

Improved Heat Map