Take-Home Exam

Data Science: Visualizations & Analytics in R (S_VAR) - 6 ECTS

dr. Mariken A.C.G. van der Velden

December 13, 2019

Instructions for the take-home exam

This exam is available from Tuesday, 13 December, at 9am. You can hand in your answers until Thursday, 19 December, 23.59h via Canvas. Write down your answers - including R code for the visualizations - with **the same number-letter combination** in a separate document (e.g., Google Doc), and submit this document to Canvas in a PDF format.

The exam consists of two tasks: (1) Develop an infographic; and (2) Commenting on the infographic. Each task accounts for 50% of the exam grade. The exam counts for 60% of your final grade.

1. Develop an Infographic

Use the datasets `economics`, `midwest` (both available in `ggplot2`). Based on these datasets, you are going to develop **1** infographic that demonstrates the following sub-graphs:

1. Show how the unemployment rate in the Midwest (US) has changed since the 1960s and especially since the beginning of the financial crisis.

2. Compare unemployment rate with median length of unemployment in one graph.

3. Show the relationship between unemployment rate and duration of unemployment per decade.

4. Visualize the percentage of adults living under the poverty line in the US on a map.

1.1 Change in Unemployment Rate

1.1.a) Use the following code to generate a new dataset (`econ`) that includes a variable for unemployment rate and a dummy variable that separates the time before the beginning of the financial crisis from the period after the crisis based on the `economics` data.

```
library(tidyverse)
econ <- economics %>%
  mutate(unemployment_rate = unemploy/pop,
         crisis = ifelse(date < as.Date("2008-10-01"),
                         "before", "after" ))
```

1.1.b) Generate a plot with the following elements:

- Draw a line plot that shows the unemployment rate (Y-axis) over time (X-axis).

- Display years in 5- year steps on the x-axis using `scale_x_date()`.

- Add a colored vertical line at October 2008 using `geom_vline()`.

- Add the text "financial crisis" to the vertical line using `annotate()`:

- Remove the title on x-axis.

- Set the Y-axis to percentages.

- Use `ggplot`'s classic inbuilt theme.

1.2 Compare Unemployment Rate

1.2.a) Use the following code to add a new variable `decade` to the `econ` dataset:

```
econ <- econ %>%
  mutate(decade =
           factor(paste0(substr(date,1,3),0,
                         "-",substr(date, 1,3),9)) )
```

1.2.b) Generate a plot with the following elements:

- Make a scatterplot with `unemployment_rate` on the X-axis and the length of unemployment in weeks on the Y-axis, color the points with `firebrick1`. Tip: check `?economics` for an explanation of the variables in the data.

- Use `facet_wrap()` to draw one plot per decade. Use `ncol = 2`.

- Add a black regression line with `geom_smooth()`.

- Set the X-axis to percentages.

- Use `ggplot`'s classic inbuild theme.

1.3 Relationship between Rate and Duration of Unemployment

1.3.a) Use the following code to run a linear regression with the duration of unemployment as dependent variable and decade as independent variable (removing the intercept). This code saves the regression results (coefficients and confidence intervals) in a data frame.

```
m1 <- lm(uempmed ~ 0 + decade, data = econ)

results <- data.frame(decade =
                      gsub("decade","",rownames(confint(m1))),
                        pred = coef(m1),
                        lower = confint(m1)[,1],
                        upper = confint(m1)[,2])
```

1.3.b) Generate a plot with the following elements:

- Make a scatterplot with the decade on the X-axis and the predicted duration on the Y-axis, color the points with `firebrick1` and set their size to 3.

- Add an error-bar around the estimations using `geom_errorbar()`.

- Flip the X and Y-axis.

- Use `ggplot`'s classic inbuild theme.

1.4 Geographical Mapping

In the new version of the take home exam, it's specified that you have to fill the statebins with percentages of adults living under poverty line (percadultpoverty in midwest).

1.4.a) Use the `midwest` dataset.

1.4.b) Generate a plot with the following elements:

- Make a statebin graph using `statebins_continuous`. Don't forget to call the `statebins` library.

- Make the text in the graph white and use font size 10.

- Color the bins red.

- Hide the legend.

1.4.c) Combine the four graphs with `multiplot()`. Make sure the code for the four graphs as well as the final figure is copied into your take-home exam.

2. Comment on Your Own Infographic

Based on the above indicated steps, your infographic looks similar to Figure 1 (see below). Use approx. 300 words to comment on your infographic, addressing the following questions:

1. What do each of the individual graphs demonstrate: write it up as if you would inform an *Economist* reader.

2. What if you would write for a tabloid newspaper, would you assess the visual literacy of the readership to change? If so, what would you adjust in the infographic?

3. What are the strengths of this infographic, what are the weaknesses? Would you recommend a data scientists writing a report for a general audience to use it, and if so, for which purposes, and if not, why?

Good Luck

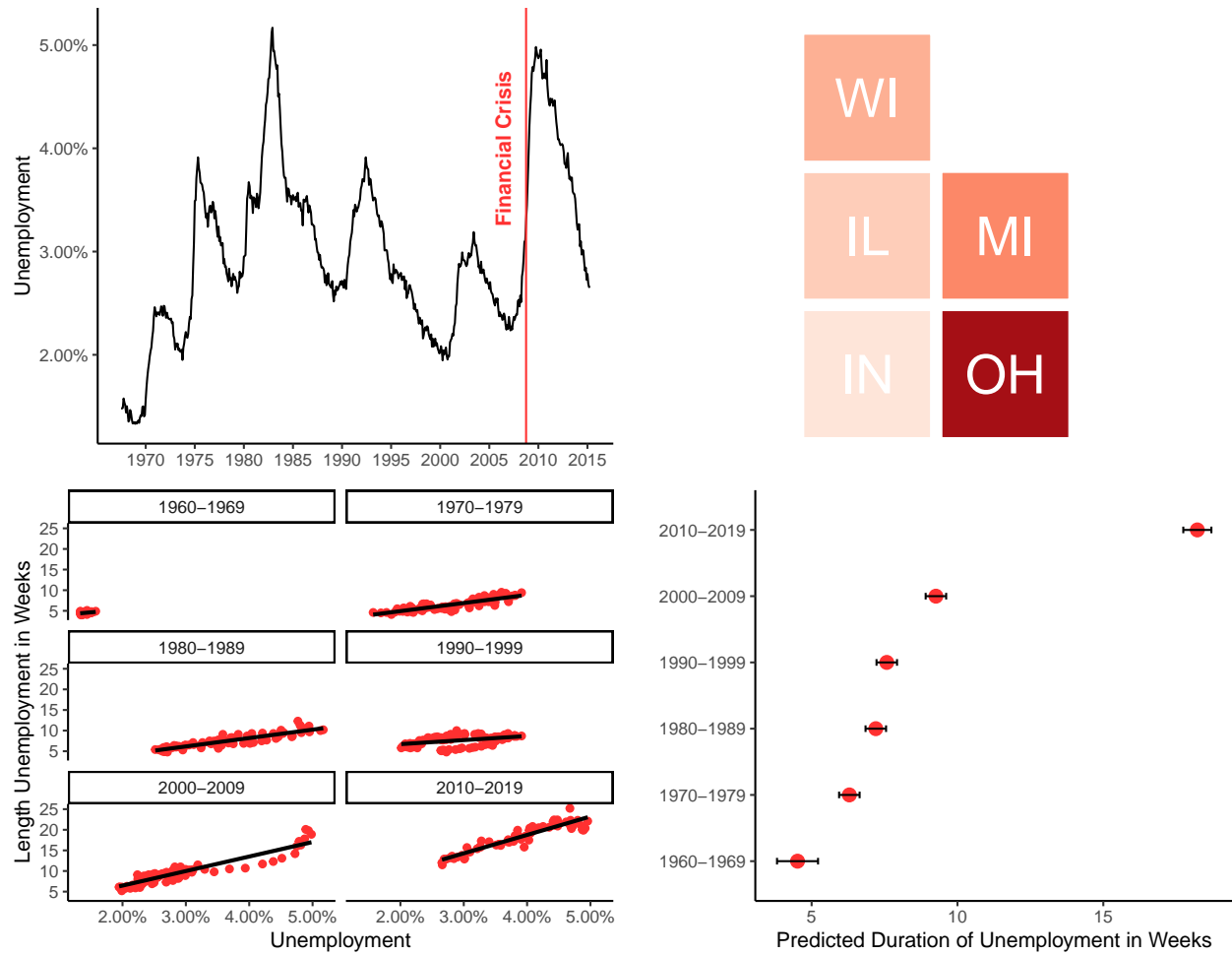The deadline for the take-home exam is **December 19th, midnight**!

Figure 1: The Financial Crisis and Its Effect in the Midwest.