Lab No. 4

Elizabeth Walter

27 September, 2021

```
knitr::opts_chunk$set(echo = TRUE,
                      fig.align = 'center')
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.1.1
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.3     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1
```

```
## Warning: package 'tidyr' was built under R version 4.1.1
```

```
## Warning: package 'readr' was built under R version 4.1.1
```

```
## Warning: package 'forcats' was built under R version 4.1.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(geofacet)
```

```
## Warning: package 'geofacet' was built under R version 4.1.1
```

```
unemployment_data <- read.csv("unemployment.csv")
```

Exercise

Use data from the US Bureau of Labor Statistics (BLS) to show the trends in employment rate for all 50 states between 2006 and 2016.

What stories does this plot tell? Which states struggled to recover from the 2008–09 recession?

```r
unemp_small <- unemployment_data %>%
  mutate(state = reorder(state, unemployment))

unemploy_plot <- ggplot(unemp_small, aes(x = date, y = unemployment)) +
  geom_point() +
  facet_wrap(vars(state)) +
  theme_minimal() +
  labs(x = NULL,
       y = "Unemployment Rate (%)",
       title = "Unemployement Rates for US States & D.C., 2006 - 2016") +
  theme(axis.text.x = element_blank(),
        panel.grid.minor.x = element_blank())

unemploy_plot
```

Unemployement Rates for US States & D.C., 2006 – 2016
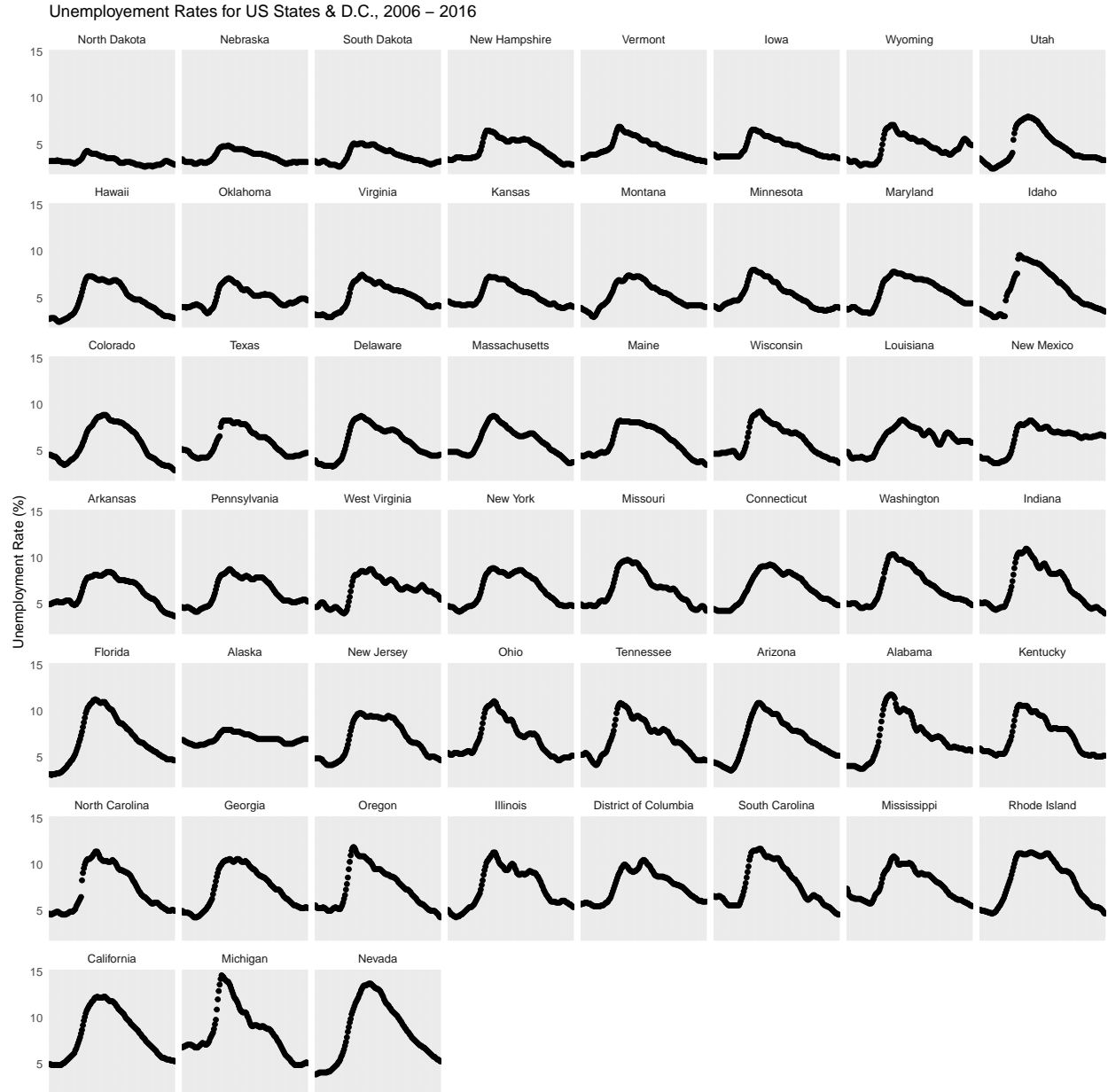
I reordered the data to display the plots by lowest to highest unemployment, based on the state's highest unemployment rate across the time period. From this we see that the states that saw the highest unemployment rates during the recession are not from a particular region or division - the four states with highest peaks in unemployment are Rhode Island, California, Michigan, and Nevada. We also see differing trends of recovery, as some states saw generally smooth declines from their peaks, such as Florida, California, and Nevada, while other states saw spikes (ex: New York, Indiana, Massachusetts) or plateaus (ex: Missouri, Michigan, Kentucky) in their recovery. Many states saw unemployment rates in 2016 near their 2009 rates, with a few even seeing lower rates at the end of the period. A few notable exceptions are New Mexico, Louisiana, Alabama and Wyoming, as the difference between the 2019 and 2006 unemployment rates in those states were much greater, relative to most other states.

Slopegraphs

Use data from the BLS to create a slopegraph that compares the unemployment rate in January 2006 with the unemployment rate in January 2009, either for all 50 states at once (good luck with that!) or for a specific region or division. Make sure the plot doesn't look too busy or crowded in the end.

What story does this plot tell? Which states in the US (or in the specific region you selected) were the most/least affected the Great Recession?

```r
unemp_06_09 <- unemployment_data %>%
  filter(year %in% c(2006, 2009)) %>%
  filter(month == 1) %>%
  mutate(year = factor(year)) %>%
  group_by(year, division) %>%
  summarize(avg_unemp = mean(unemployment)) %>%
  mutate(label_first = ifelse(year == 2006, paste0(division, ": ", round(avg_unemp, 2), ' %'), NA),
         label_last = ifelse(year == 2009,  paste0(round(avg_unemp, 2),' %'), NA))
```

```
## 'summarise()' has grouped output by 'year'. You can override using the '.groups' argument.
```

```r
slopegraph_06_09 <- ggplot(unemp_06_09, aes(year, avg_unemp, group = division)) +
  geom_line(aes(color = division)) +
  theme_bw() +
  theme(legend.position = 'none') +
  geom_text(aes(label = label_first, color = division), direction = 'y', nudge_x = -0.3, show.legend = 
  geom_text(aes(label = label_last, color = division), direction = 'y', nudge_x = 0.1, show.legend = FA
  xlab("") +
  ylab("Unemployment Rate") +
  ggtitle("Change in Unemployment Rate by Division, 2006 vs. 2009")
```
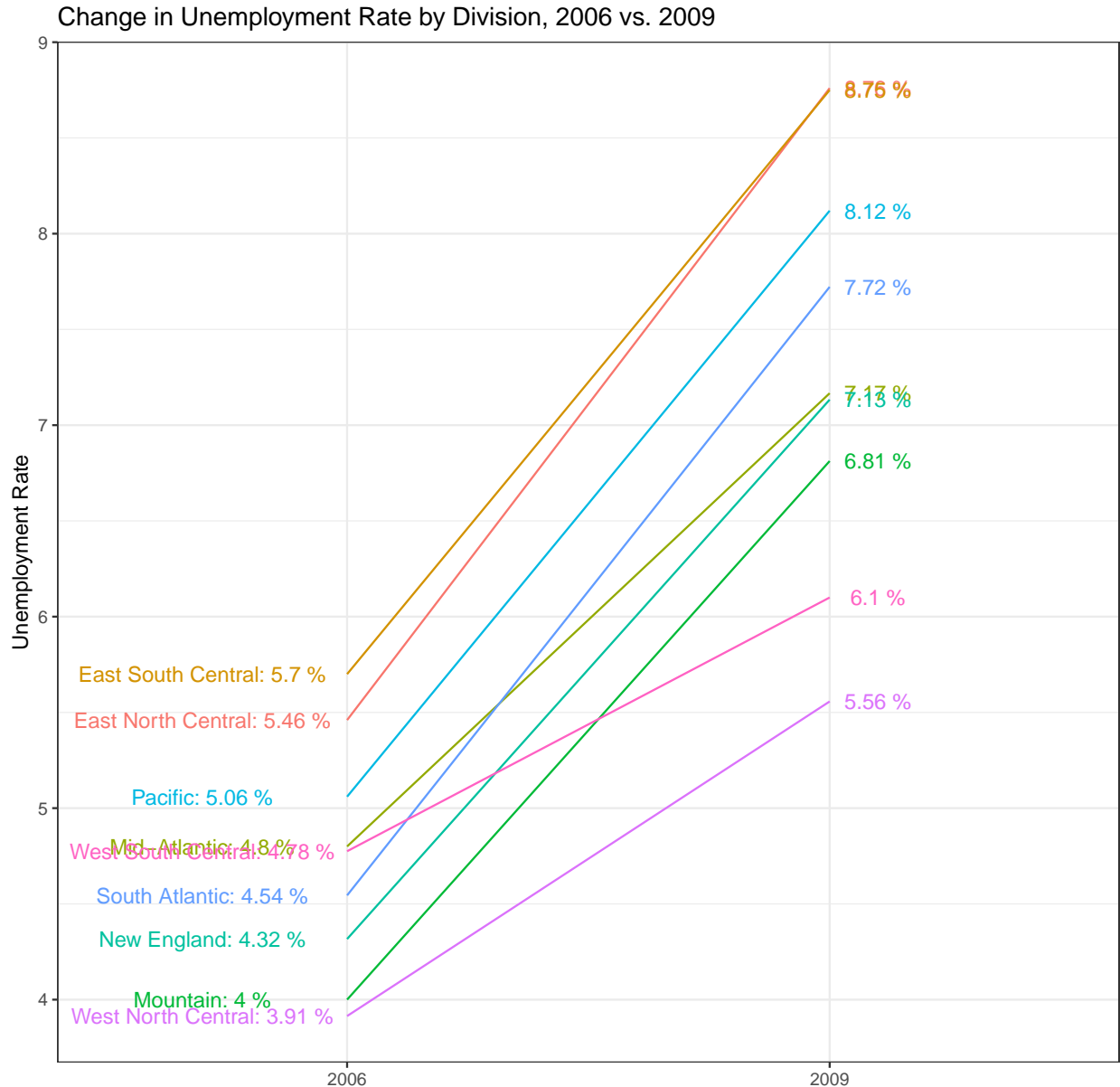
```
## Warning: Ignoring unknown parameters: direction
```

```
## Warning: Ignoring unknown parameters: direction
```

```r
slopegraph_06_09
```

```
## Warning: Removed 9 rows containing missing values (geom_text).
```

```
## Warning: Removed 9 rows containing missing values (geom_text).
```

Change in Unemployment Rate by Division, 2006 vs. 2009

I grouped by division, which showed that with a few exceptions, changes in unemployment from 2006 to 2009 were very similar across divisions, and rankings of divisions by unemployment rate remained the same from 2006 to 2009, with again, a few exceptions. The exceptions: West North Central (Missouri, Kansas, Nebraska, Iowa, North Dakota, South Dakota, Minnesota) maintained its rank as lowest unemployment rank in each year, but it also saw a smaller increase in the unemployment rate, as did West South Central (Arkansas, Oklahoma, Texas, Louisiana), which moved it from the fifth lowest to second lowest unemployment rank from 2006 to 2009. The last exception that also saw a smaller increase in unemployment and changed rank across the period was the Mid-Atlantic division (New Jersey, New York, Pennsylvania). The East South Central (Alabama, Mississippi, Tennessee, Kentucky) and East North Central (Wisconsin, Illinois, Indiana, Ohio, Michigan) divisions had the highest unemployment rates before the recession in 2006, but also had the highest rates in 2009, with the East North Central division appearing to have the greatest increase, among all divisions, in unemployment between the two years.