Lab 03 - Tidy Graphics and Teamwork

Due: Sunday, Feb 2 at 11:59pm

Packages

In this lab we will work with the `tidyverse` package. Load the package with

```
library(tidyverse)
```

```
## -- Attaching packages -------------------------- tidyverse 1.3.0 --

## v ggplot2 3.2.1     v purrr   0.3.3
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0

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

Data

In the **Files** pane you will see a folder named `data`. In it are three CSV files. Your team will work with data in one of these files according to the following rule:

Lab Section	Data
04: 1:25p - 2:40p	`fishing.csv`
05: 3:05p - 4:20p	`employees.csv`
06: 4:40p - 5:55p	`nerrs.csv`

CSV files can be read into R using function `read_csv()`. In the function we'll need to specify where the file is located. We will also want to save the result so we can work with the data in subsequent steps. For example, if we had a file named `coronavirus.csv` in the `data` folder, I would read it into R with

```
coronavirus <- read_csv("data/coronavirus.csv")
```

1. Assign each person on your team a number 1 through 4. For teams of three, Member 1 can take on the role of Member 4 in Exercise 5.

2. **Member 1**: Team gitdata: Nathan Kim, Avanti Shah, Blossom Mojekwu.

```
employees <- read_csv("data/employees.csv")
```

```
## Parsed with column specification:
## cols(
##   employee_type = col_character(),
##   year = col_double(),
##   percent = col_double()
```

```
## )
```

3. **Member 2**:

```
ggplot(data = employees, mapping = aes(x = year, y = percent, color = employee_type)) +
    geom_line() + labs(title = "Employment rate for different employee types (1975-2011)", x = "Year",
    y = "Employment rate (%)") + theme(legend.position="right",  legend.title = element_text(size = 10),
```

Employment rate for different employee types (1975–2011)
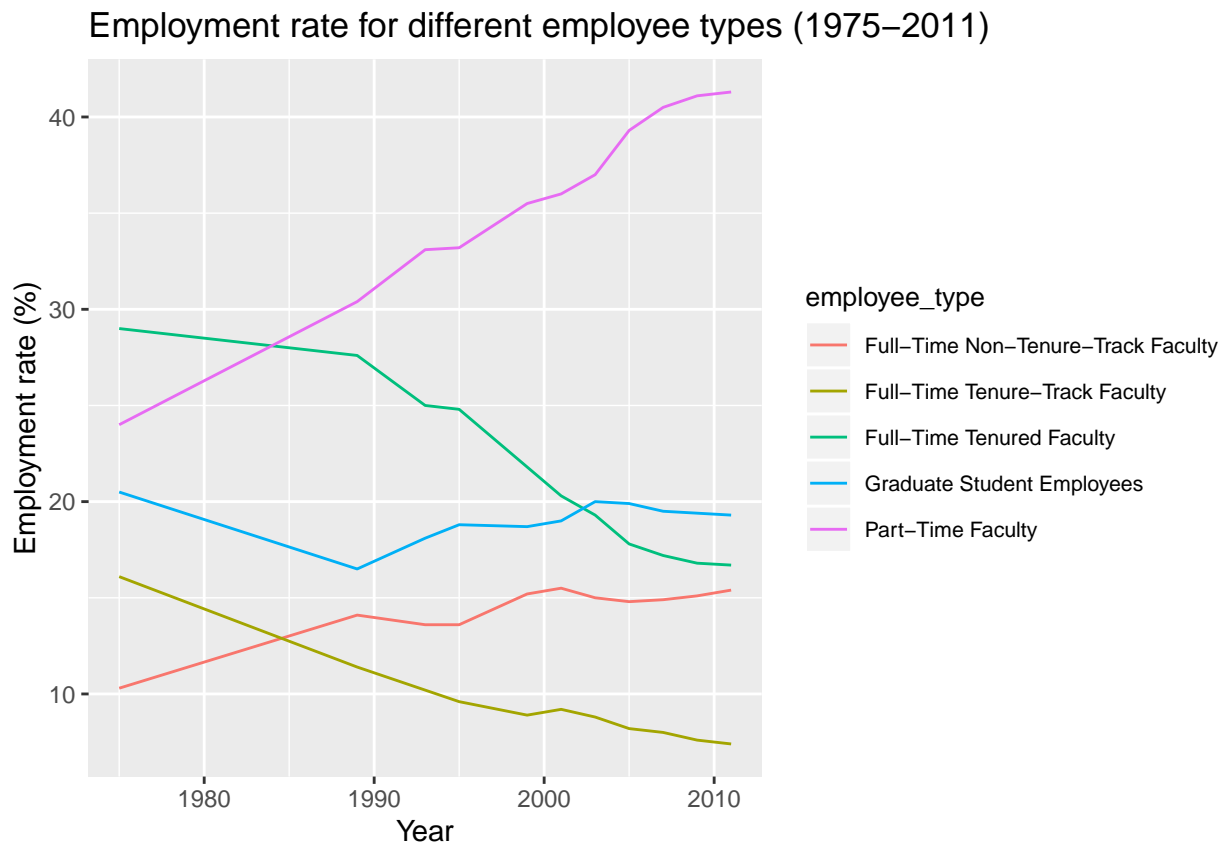
4. **Member 3**:
   The original graphic was poor because it made readibility harder for the viewer. The x axis contained the percentage while the y axis contained the the employee type. Generally, numeric values are placed on the y-axis. The original graph also had too many colors making it distracting because each color was associated with a year and grouped in bars by employee-type making it hard to distinguish the data. The background was also a similar color with the colors in the legend. The different colors for the year didn't allow the reader to see the percentage change throughout the years. The graph also had too many bars to display the percentages per year.

5. **Member 4**: Our graphic is an improvement over the original graphic in many ways. First, we've changed the axes and legends in order to better portray change across time. This was done in conjunction with changing the graphic from a bar graph to a line graph. Line graphs are typically better at showing continuous change over time. The use of a grey background also made some of the colors(shades of blue) stand out more. The new graph is also completely 2D, whereas the original graph was in 3D, with the bars being gradient and the edges of the graph having shading, which are both potentially distracting.