Data Viz.: Homework 1

Naimul Chowdhury

May 18, 2020

Reading:

a. Chapters 1 and 2 in Kieran Healy's book "Data Visualization. Chapter 2 should mostly be familiar material.
b. Read my R handout on Exploring the GSS, which I will add to after class on Thursday.

##1. Setup ### options Set up global options

libraries

Load in needed libraries

```
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   1.0.2     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()
```

```
library(RColorBrewer)
library(haven)
```

2. File management

Create variables for directories

```
project.dir <- getwd() #naeem
output.dir <- "/Output"
data.dir <- "C:/Users/Naeem Cho/Desktop/School Work/Data_Viz/Datasets"
setwd(project.dir)
getwd()
```

```
## [1] "C:/Users/Naeem Cho/Desktop/School Work/Data_Viz/Data_Viz/2020-02-04_HW1"
```

3. Importing Data

```
GSS_Data <- read_dta(file.path(data.dir, "GSS2018.dta"))
```

Project:

1. Pick out a **categorical** variable (i.e., Question) from the General Social Survey and visualize its distribution. Think carefully about elements of your display (title, subtitle, choice of colors, axis labels, displaying or not displaying percentages with the bars, etc.) so that it tells a clear story. Note that there is no right answer, although there certainly can be bad displays. In addition, prepare three slides that summarize the key steps you have taken in R to arrive at your final display. You may have to present these slides in class on Thursday next week. Make sure your slides use a large enough font size for all of us to see what you were doing and for us to get the general idea of the steps you have taken.

I'm most interested in the *othlang* and *abfelegl* categorical variables.

Other variables of interest: *othlang, othlang2, livewhts, owngun, homosex, prayer, jew, relig, spkmslm, spkhomo, colcom, spkrac, colrac*

Can the respondent speak language other than English/Spanish?

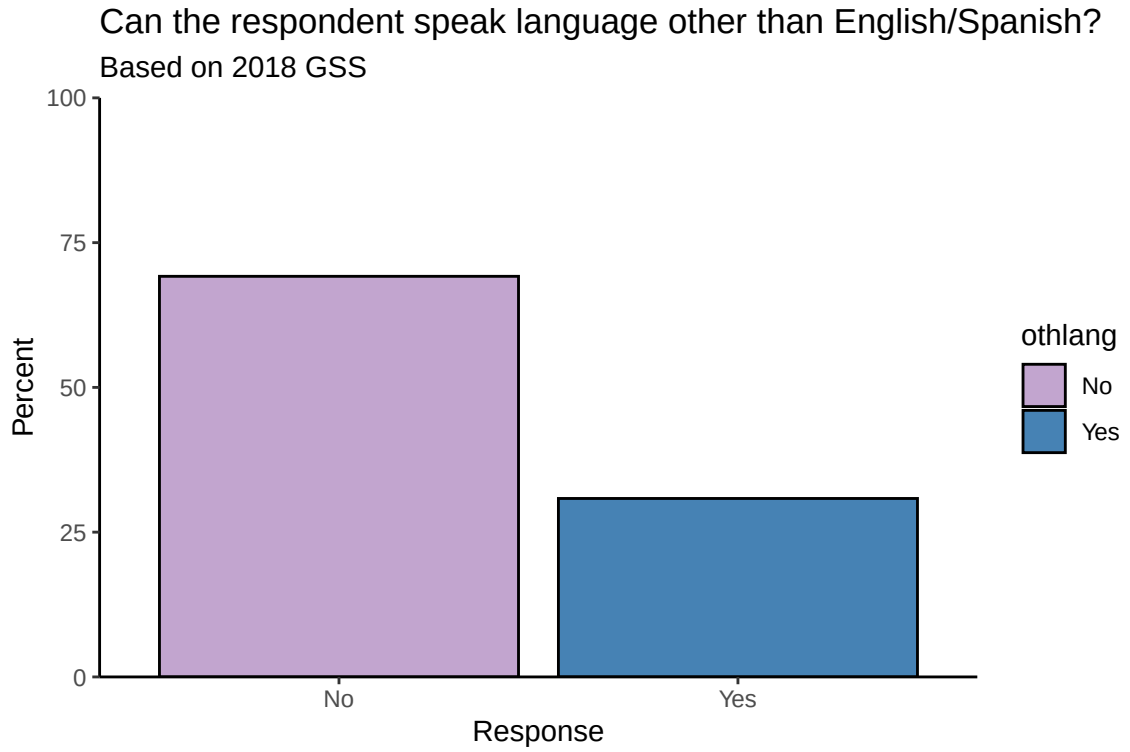
```
#The Color Palette
mycols <- c(brewer.pal(11, "PRGn")[4], "steelblue", brewer.pal(11, "PRGn")[8])

othlang <- GSS_Data %>%
  select(othlang)

props <- 100* prop.table(table(othlang))
props
```

```
## othlang
##        1        2
## 30.81841 69.18159
```

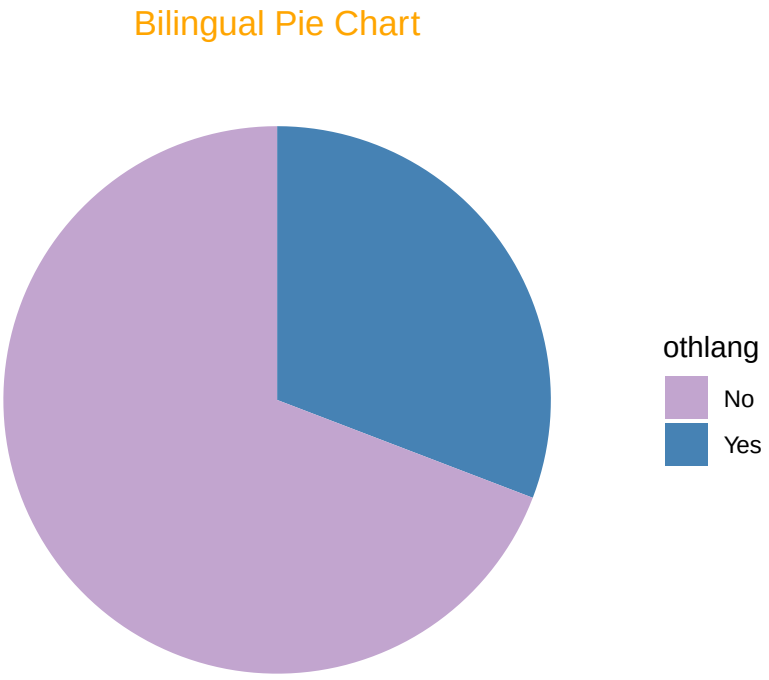
```
lang_table <- data.frame(othlang = c("Yes", "No"), props = as.vector(props))
str(lang_table)
```

```
## 'data.frame':    2 obs. of  2 variables:
##  $ othlang: Factor w/ 2 levels "No","Yes": 2 1
##  $ props  : num  30.8 69.2
```

```
ggplot(data= lang_table, aes(x=reorder(othlang,-props),y=props, fill=othlang)) +
  geom_bar(stat="identity", color = "black") +
  scale_fill_manual(values=mycols) +
  theme_classic() + ylab("Percent") + xlab("Response") +
  scale_y_continuous(limit=c(0,100), expand=c(0,0)) +
  ggtitle(label="Can the respondent speak language other than English/Spanish?", subtitle="Based on 201
```




```
ggplot(data = lang_table, aes(y = props, fill = othlang)) +
  geom_bar(aes(x = "1"), stat = "identity") +
  coord_polar("y") +
  scale_fill_manual(values=mycols) +
  theme_classic() +
  theme(axis.line = element_blank(),
        axis.ticks = element_blank(),
        axis.text = element_blank(),
        axis.title = element_blank(),
        plot.title = element_text(hjust = 0.5, color = "orange")
        ) +
  ggtitle(label = "Bilingual Pie Chart")
```



Turn in (upload) one pdf knitted from your R document which shows how you obtained both of the displays.
Turn in (upload) the 6 slides you have created.