# Homework 4

*Angela Zhang*

*10/23/2017*

## The Rules

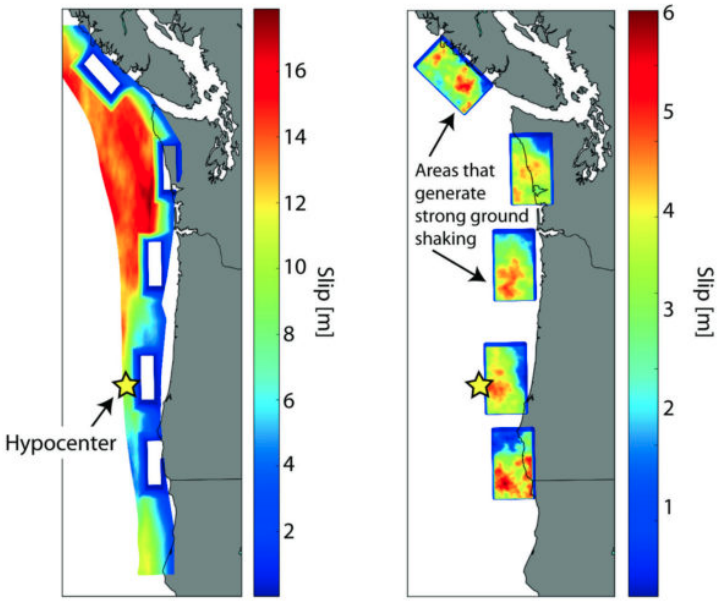



Figure 1: Simulation of the expected Seattle earthquake

(1)

This figure utilizes color (Rule 9) well since it depicts the greatest slippage with an intense color (red). Labelling could be improved in this figure (Rule 5). While they show where the hypocenter of the earthquake is, I'm not quite sure where Seattle is or what states (Washington and Oregon perhaps?) are depicted in the figure. This figure could also be improved if distance, latitude, and longitude were added to the map.

(2)

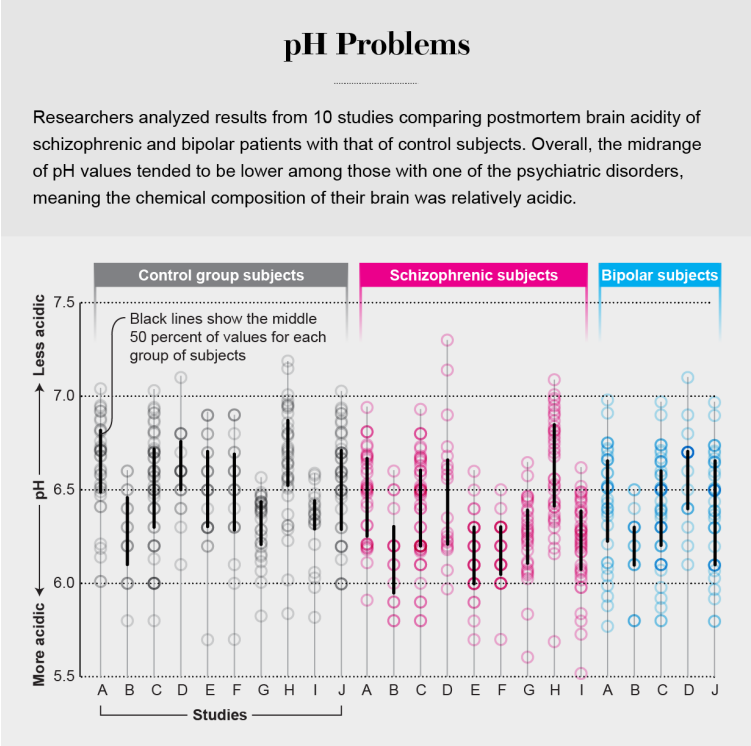



Figure 2: Article from Scientific American

This plot use colors (Rule 9) and labels (Rule 5) well for the most part. The colors indicate the different groups and the artist uses transparencies well to show where points cluster together. The short description at the top is very helpful for understanding what figure is trying to show. It's a bit of a complicated graph to read and the results don't seem quite apparent at first, but the labels are a good aid for understanding the x-axis labels and what the black lines in the graph. The description, as helpful as it is, may bias the reader. It states that schizophrenic patients have a more acidic brain environment, but this pattern is not clear in the chart.

(3)

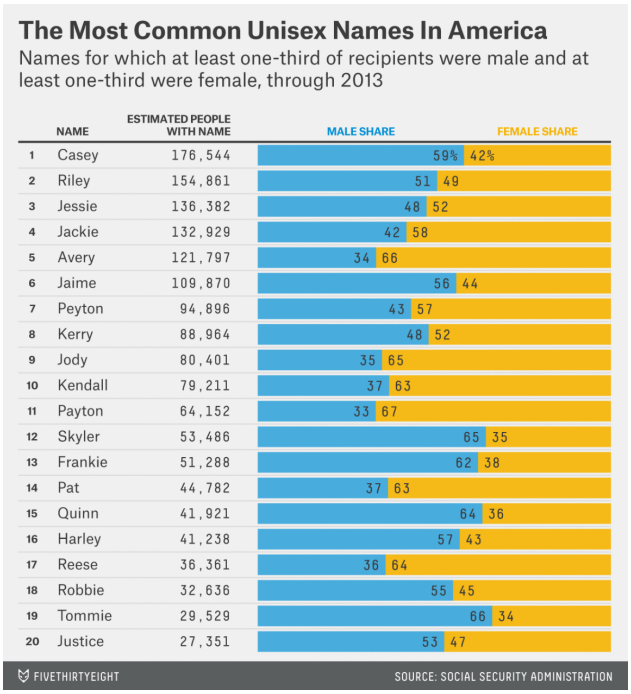
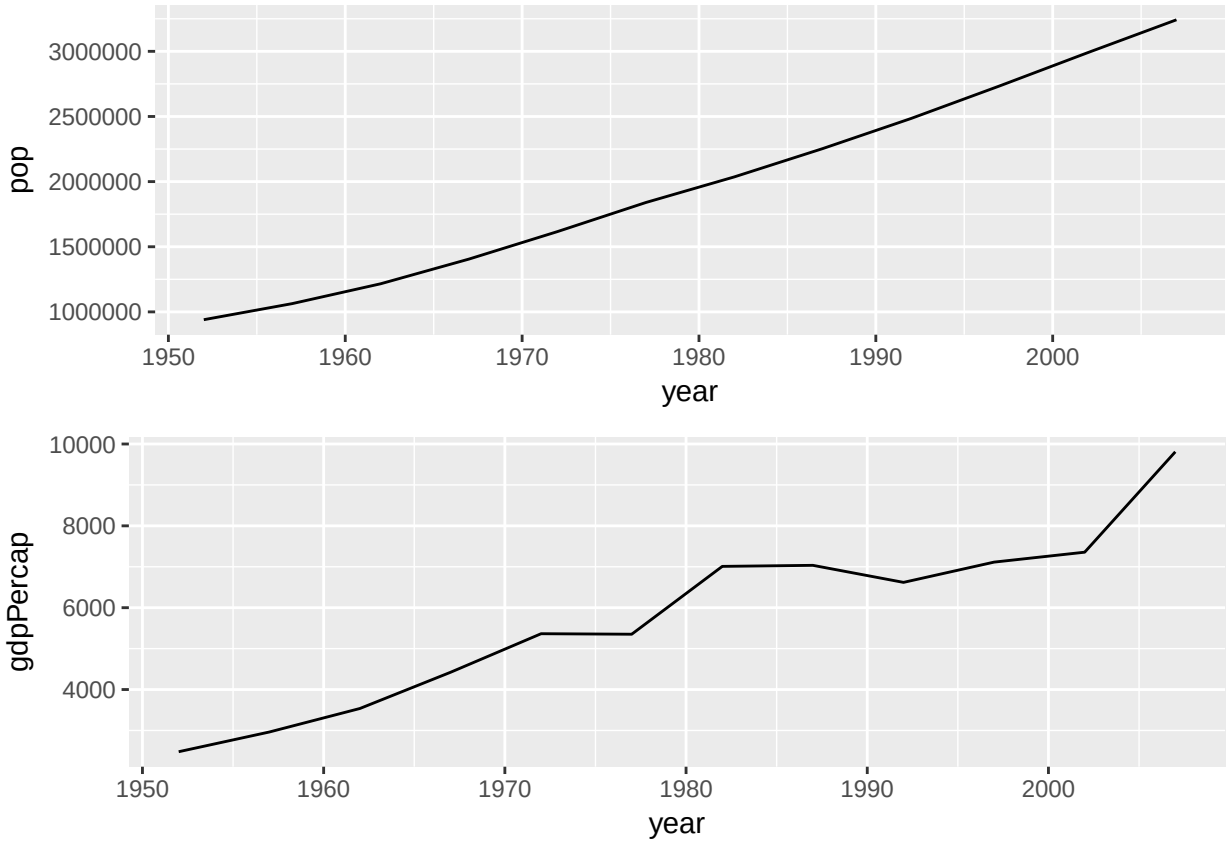



Figure 3: Chart from FiveThirtyEight

This figure does a good job of utilizing color and labels (Rule 9 and Rule 5 respectively) as it explains the proportion of females and males with unisex names. It it's pleasing to the eye while avoiding chart junk (Rule 3). It also incorporates aspects of Rule 1 by setting the table side by side with the stacked bar plots. Overall, I don't have many criticisms for this figure since it seems to follow many of Tufte's Rules.

### Multiple Plots




I plotted these graphs on top of each other so that the x-axes matched together. That way, you can view the trend of both population and GDP as year increases.

### Data wrangling with dates and legends

```
library(readr)
library(reshape)
abund = read.delim("~/Documents/Biost561/Materials/biostat561/lecture4/abundances.txt",
                   sep = "\t",row.names = 1, check.names= F)

sample = abund[c(sample(1:448, 6, replace=F)),]
taxon = as_tibble(cbind(date = names(sample), t(sample)))

taxon %>%
  separate(col = date, into = c("Day","Month","Year","Extra"), sep = "\\.") %>%
  select(-Extra) -> taxon1
```

```
## Warning: Too few values at 4 locations: 2, 3, 4, 70
```

```
taxon1$Year[nchar(taxon1$Year) == 2] = paste(20,taxon1$Year[nchar(taxon1$Year) == 2], sep="")

taxon1 %>%
  unite(col = Date, c(Day,Month,Year), sep = "-") %>%
  group_by(Date) %>%
```

```
  mutate_if(is.character, as.numeric) %>%
  summarise_at(.vars = names(.)[2:7],
               .funs = c(mean="mean")) -> mean_taxon

mean_taxon = mean_taxon[rowSums(mean_taxon[2:7])>0,]
melt_taxon = melt(as.data.frame(mean_taxon), id="Date")
melt_taxon = melt_taxon[order(as.Date(melt_taxon$Date, format="%d-%m-%Y")),]
melt_taxon$Date = factor(melt_taxon$Date, levels = unique(melt_taxon$Date))
```

```
## ggplot

library(ggplot2)

ggplot(melt_taxon, aes(x= Date, y = value, fill = variable)) +
  geom_bar(stat='identity', position = "fill") +
  theme(axis.text.x = element_text(size = 6,angle = 90, hjust = 1,),
        plot.title = element_text(hjust = 0.5)) +
  ggtitle("Change of taxon proportions over time") + labs(x="Date",y="Proportions")
```