

Lab 12 - Statistics, Coordinates, Facets, and Themes

Yana Chakalo

November 14, 2017

Complete the following exercises below. Knit together the PDF document and commit both the Lab 12 RMDfile and the PDF document to Git. Push the changes to GitHub so both documents are visible in your public GitHub repository.

1. Choose one or more graphics you created for Lab 11 and either experiment with the underlying statistical layer if it already has one (i.e. if you made a histogram experiment with different bin widths) or add a separate statistical layer to your plot (i.e. a smoothing curve). Choose something you think will offer meaningful insight and describe why you made the choice you did. What additional information does this provide viewers of your graphic?

I organized my data in a way that included all the categorical variables into descriptions that people will be able to read instead of the numbers that were original assigned for each category. I did this so it would not be tiresome to consistently look at the codebook because my data has a lot of variables.

BIASMO1 histograms, 2002 & 2013

```
theme_set(theme_classic())
```

```
## Error in theme_set(theme_classic()): could not find function "theme_set"
```

```
g <- ggplot(data = data_subset2, aes(x=as.factor(BIASMO1)))
```

```
## Error in ggplot(data = data_subset2, aes(x = as.factor(BIASMO1))): could not find function "ggplot"
```

```
g + geom_bar(stat="count", width = 0.5, fill="tomato2") +  
  labs(title="Frequency of BIASMO1, 2002 Bar Chart",  
        subtitle="The rate of each Bias motivation",  
        caption="Type of Bias Motivation") +  
  theme_hc() +  
  theme(axis.text.x = element_text(angle=90, vjust=0.6)) +  
  scale_x_discrete(labels=c("Anti-White", "Anti-Black", "Anti-Am Indian", "Anti-Asian", "Anti-Multi-Racial"))
```

```
## Error in eval(expr, envir, enclos): object 'g' not found
```

```
g <- ggplot(data = data_subset13, aes(x=as.factor(BIASMO1)))
```

```
## Error in ggplot(data = data_subset13, aes(x = as.factor(BIASMO1))): could not find function "ggplot"
```

```
g + geom_bar(stat="count", width = 0.5, fill="tomato2") +  
  labs(title="Frequency of BIASMO1, 2013 Bar Chart",  
        subtitle="The rate of each Bias motivation",  
        caption="Type of Bias Motivation") +  
  theme_hc() +  
  theme(axis.text.x = element_text(angle=90, vjust=0.6)) +  
  scale_x_discrete(labels=c("Anti-White", "Anti-Black or African American", "Anti-American Indian or Alaska Native", "Anti-Asian", "Anti-Multi-Racial"))
```

```
## Error in eval(expr, envir, enclos): object 'g' not found
```

GOFFRAC histograms, 2002 & 2013

```
g <- ggplot(data = data_subset2, aes(GOFFRAC))

## Error in ggplot(data = data_subset2, aes(GOFFRAC)): could not find function "ggplot"

g + geom_bar(stat="count", width = 0.5, fill="blue") +
  labs(title="Frequency of GOFFRAC, 2002 Bar Chart",
        subtitle="The Frequency of Offenders Race",
        caption="Frequency of Offenders Race") +
  theme_hc() +
  theme(axis.text.x = element_text(angle=65, vjust=0.6)) +
  scale_x_discrete(labels=c("Asian", "Black or African American", "American Indian/Alaska Native", "Mul

## Error in eval(expr, envir, enclos): object 'g' not found

g <- ggplot(data = data_subset13, aes(GOFFRAC))

## Error in ggplot(data = data_subset13, aes(GOFFRAC)): could not find function "ggplot"

g + geom_bar(stat="count", width = 0.5, fill="blue") +
  labs(title="Frequency of GOFFRAC, 2013 Bar Chart",
        subtitle="The Frequency of Offenders Race",
        caption="Frequency of Offenders Race") +
  theme_hc() +
  theme(axis.text.x = element_text(angle=65, vjust=0.6)) +
  scale_x_discrete(labels=c("Asian", "Black or African American", "American Indian/Alaska Native", "Mul

## Error in eval(expr, envir, enclos): object 'g' not found
```

LOCCOD1 Histograms 2002 & 2013

```
g <- ggplot(data = data_subset2, aes(x=as.factor(LOCCOD1)))

## Error in ggplot(data = data_subset2, aes(x = as.factor(LOCCOD1))): could not find function "ggplot"

g + geom_bar(stat="count", width = 0.5, fill="green") +
  labs(title="Frequency of LOCCOD1, 2002 Bar Chart",
        subtitle="The Frequency of Incident Location",
        caption="Frequency of Where the Hate Crime Took Place") +
  theme_hc() +
  theme(axis.text.x = element_text(angle=90, vjust=0.6)) +
  scale_x_discrete(labels=c("Terminal", "Bank", "Bar", "Church/Synagogue/Temple", "Office building", "Con

## Error in eval(expr, envir, enclos): object 'g' not found

g <- ggplot(data = data_subset13, aes(x=as.factor(LOCCOD1)))

## Error in ggplot(data = data_subset13, aes(x = as.factor(LOCCOD1))): could not find function "ggplot"

g + geom_bar(stat="count", width = 0.5, fill="green") +
  labs(title="Frequency of LOCCOD1, 2013 Bar Chart",
        subtitle="The Frequency of Incident Location",
        caption="Frequency of Where the Hate Crime Took Place") +
  theme_hc() +
```

```

  theme(axis.text.x = element_text(angle=90, vjust=0.6)) +
  scale_x_discrete(labels=c("Terminal", "Bank", "Bar", "Church/Synagogue/Temple", "Office building", "Con

## Error in eval(expr, envir, enclos): object 'g' not found

```

STATECOD histograms, 2002 & 2013

```

g <- ggplot(data = data_subset2, aes(x=as.factor(STATECOD)))

## Error in ggplot(data = data_subset2, aes(x = as.factor(STATECOD))): could not find function "ggplot"

g + geom_bar(stat="count", width = 0.5, fill="purple") +
  labs(title="Frequency of STATECOD, 2002 Bar Chart",
        subtitle="The Frequency of States",
        caption="Frequency of States Where Hate Crimes Have Occured") +
  theme_hc() +
  theme(axis.text.x = element_text(angle=65, vjust=0.6))

## Error in eval(expr, envir, enclos): object 'g' not found

g <- ggplot(data = data_subset13, aes(x=as.factor(STATECOD)))

## Error in ggplot(data = data_subset13, aes(x = as.factor(STATECOD))): could not find function "ggplot"

g + geom_bar(stat="count", width = 0.5, fill="purple") +
  labs(title="Frequency of STATECOD, 2013 Bar Chart",
        subtitle="The Frequency of States",
        caption="Frequency of States Where Hate Crimes Have Occured") +
  theme_hc() +
  theme(axis.text.x = element_text(angle=65, vjust=0.6))

## Error in eval(expr, envir, enclos): object 'g' not found

2. With the same or a different plot created in Lab 11, experiment with zooming in on specific areas of
your graphic and changing the aspect ratio. Are there any benefits/drawbacks with either or both of
these approaches for the visualizations you've created? What are they?

```

Data from 2002. I zoomed in on two specific bias motives to compare them.

```

g <- ggplot(data = data_subset2, aes(x=as.factor(BIASM01)))

## Error in ggplot(data = data_subset2, aes(x = as.factor(BIASM01))): could not find function "ggplot"

g + geom_bar(stat="count", width = 0.5, fill="tomato2") +
  labs(title="Frequency of BIASM01, 2002 Bar Chart",
        subtitle="The rate of each Bias motivation",
        caption="Type of Bias Motivation") +
  theme(axis.text.x = element_text(angle=90, vjust=0.6)) +
  scale_x_discrete(limit = c("11", "12"), labels=c("Anti-White", "Anti-Black"))

## Error in eval(expr, envir, enclos): object 'g' not found

```

3. Try facetting a plot you have made by another categorical variable in your data (this can even be as simple as Male/Female). What is the difference between `facet_wrap()` and `facet_grid()`? How might facetting be useful in data visualization?

Facet wrap strings up different frames based on a single variable. Facet grid takes two variables like X and Y.

Facet of 2002 data. The graph shows that the population above the mean population, 1, and population below the mean 0 and its location of the hate crime incidents. We see that the cities below the population mean have higher spikes in hate crimes committed in the home.

```
mean.pop1 <- mean(data_subset2$POP1)

## Error in mean(data_subset2$POP1): object 'data_subset2' not found

data_subset2.2 <- data_subset2 %>%
  group_by(CITY) %>%
  mutate(pop.new = ifelse(POP1 >= mean.pop1, 1, 0))

## Error in data_subset2 %>% group_by(CITY) %>% mutate(pop.new = ifelse(POP1 >= : could not find function
levels(data_subset2.2$pop.new) <- c("Below Mean Pop", "Above Mean Pop")

## Error in levels(data_subset2.2$pop.new) <- c("Below Mean Pop", "Above Mean Pop"): object 'data_subse
levels(data_subset2.2$pop.new)[levels(data_subset2.2$pop.new) == "Below Mean Pop"] <- "Below Mean Pop"

## Error in levels(data_subset2.2$pop.new)[levels(data_subset2.2$pop.new) == : object 'data_subset2.2' n
levels(data_subset2.2$pop.new)[levels(data_subset2.2$pop.new) == "Above Mean Pop"] <- "Above Mean Pop"

## Error in levels(data_subset2.2$pop.new)[levels(data_subset2.2$pop.new) == : object 'data_subset2.2' n
g <- ggplot(data = data_subset2.2, aes(x=as.factor(LOCCOD1)))

## Error in ggplot(data = data_subset2.2, aes(x = as.factor(LOCCOD1))): could not find function "ggplot
g + geom_bar(stat="count", width = 0.5, fill="green") +
  labs(title="Frequency of LOCCOD1, 2002 Bar Chart",
        subtitle="The Frequency of Incident Location",
        caption="Frequency of Incident Location") +
  theme_hc() +
  theme(axis.text.x = element_text(angle=90, vjust=0.6)) +
scale_x_discrete(labels=c("Terminal", "Bank", "Bar", "Church/Synagogue/Temple", "Office building", "Con
  facet_grid(pop.new ~ .,)
```

```
## Error in eval(expr, envir, enclos): object 'g' not found
#facet_grid(~ pop.new)
```

4. Use the `theme()` layer to change the appearance of a plot of your choice including the

- plot, axes, and legend titles
- axes tick marks
- text size

- legend position
5. Create three versions of a graphic of your choice using different built-in themes or a theme created from `ggthemes`. Which ones do you think are best for presenting in an academic journal? A poster session? What are the qualities of the themes that you choose that you think make them more appropriate for presentation?

The best themes I have found that are the best to use for posters are ones that have the simplest visually but also have vertical lines to visibly show and keep track of where the Y counts are on the plot. For academic journals that best can be ones with thicker fonts and lines so everything sticks out like the theme `economist`. For poster session I think is the few theme, something that is simple and plain but has thick enough lines and tick marks to show where the variable and axis are. Overall for a journal or newspaper article I think you can use thicker lines but in presentations its best to use thinner and simpler graphics.

Trying out theme `hc` which i think is the best theme because it shows my data in the most clear of terms. (Code taken from question #1)

```
theme_set(theme_classic())
```

```
## Error in theme_set(theme_classic()): could not find function "theme_set"
```

```
g <- ggplot(data = data_subset2, aes(x=as.factor(BIASM01)))
```

```
## Error in ggplot(data = data_subset2, aes(x = as.factor(BIASM01))): could not find function "ggplot"
```

```
g + geom_bar(stat="count", width = 0.5, fill="tomato2") +
  labs(title="Frequency of BIASM01, 2002 Bar Chart",
        subtitle="The rate of each Bias motivation",
        caption="Type of Bias Motivation") +
  theme_hc() +
  theme(axis.text.x = element_text(angle=90, vjust=0.6)) +
  scale_x_discrete(labels=c("Anti-White", "Anti-Black", "Anti-Am Indian", "Anti-Asian", "Anti-Multi-Racial"))
```

```
## Error in eval(expr, envir, enclos): object 'g' not found
```