

Jeffrey_exercise5_hw1

Jack Jeffrey

2024-09-10

Exercise 5

clean up workspace

```
rm(list=ls(all=TRUE)) # remove all objects
cat("\014") # clean console
```

load data

```
setwd("~/Documents/Poli502_Jeffrey/hw1/Data") # set working directory
world_data <- read.csv("~/Documents/Poli502_Je
head(world_data) # data effectively displayed
```

democracy and female representation

```
summary(world_data) #investigate the data view (world_data)
library(tidyverse) #installed tidyverse
colnames(world_data) # checking column names
```

create a scatterplot to evaluate varibales

```
ggplot(drop_na(world_data, women09), aes(x = democ_regime, y = women09, color = democ_regime)) +
geom_jitter(width = 0.2, height = 0) + labs(title = "Scatterplot of Female Representation by Regime", x =
"regime", y = "female representation") + scale_color_manual(values = c("Yes" = "green", "No" = "red"))
# scatterplot effectively created with NA's removed
```

create a boxplot

```
ggplot(drop_na(world_data, women09), aes(x = democ_regime, y = women09, color = democ_regime)) +
geom_boxplot(width = 0.4) + labs(title = "Scatterplot of Female Representation by Regime", x = "regime",
y = "female representation") + scale_color_manual(values = c("Yes" = "green", "No" = "red")) # created
a boxplot to evaluate female representation and democracy vs. non democracy
```

initial data clean up

analyze y variable

```
summary(world_data $ women09) # analyze x variable summary(world_data $ democ_regime)
###create a smaller data set that omits NA observations
women_dem <- world_data [ is.na(world_data $ women09) == FALSE & is.na(world_data $ democ_regime)
== FALSE, ]
```

```
summary(women_dem $ women09) # no observed NA's summary(women_dem $ democ_regime)
```

Re-labeling the values

```
women_dem $ dem.dum <- factor(women_dem $ democ_regime, levels = c("Yes", "No"), labels = c("Democracy", "Autocracy"))
```

Q1. Describe Y

```
summary(women_dem $ women09)
```

numeric summary for Y: The mean at 17.23 seems quite low and gives an indication that there is not strong female representation in governments across the world in 2009. It is also interesting to note the minimum which is at 0.00 indicating there is at least one country in which there is no female representation. The max at 56.30 indicates there is at least one country which has over 50% female representation.

re-graphing with adjusted data for just Y

```
ggplot(women_dem, aes(x = women09)) + geom_histogram(binwidth = 1, color = "black", fill = "blue") + labs(title = "Histogram of Female Representation", x = "representation", y = "number of observations")
```

graphical summary for Y: Using the histogram to graph Y it is evident that the majority of regimes do not have over 30% female representation. Additionally, there are almost 10 regimes which have 0 female representation. The one outlier at 56.30 can also be spotted on the histogram.

Q2. Describe X

```
table(women_dem$democ_regime) summary(women_dem $ women09) # numeric summary for X: There is not as much to analyze with this variable because it is discrete. We can see though that there are a total of 178 regime types counted. The class and mode are both character. Using the table we can see that there are 111 democracies and 67 autocracies.
```

re-graphing with adjusted data for just X

re-graphing with adjusted data for just Y

```
ggplot(women_dem, aes(x = democ_regime, fill = democ_regime)) + geom_bar(color = "black") + labs(title = "bar chart regime type", x = "regime type", y = "count") + scale_fill_manual(values = c("Democracy" = "green", "Autocracy" = "red"))
```

graphical summary for X: Using the bar chart effectively shows that there are more democracies than autocracies. There are slightly over 100 democratic regimes and slightly more than 60 autocracies.

Q3. X-Y Analysis

histogram

```
ggplot(women_dem, aes(x = women09, fill = democ_regime)) + geom_histogram(position = "identity",  
alpha = 0.5, binwidth = 1) + labs(title = "histogram of female representation by regime", x = "representation",  
y = "count") + scale_fill_manual(values = c("Yes" = "green", "No" = "red")) # this histogram effectively  
disperses the autocracies and democracies and their representation. It is not the easiest to interpret because  
certain bars contain both autocracies and democracies.
```

boxplot

```
ggplot(women_dem, aes(x = democ_regime, y = women09, fill = democ_regime)) + geom_boxplot(alpha =  
0.5) + labs(title = "boxplot of female representation by regime", x = "regime", y = "female representation")  
+ scale_fill_manual(values = c("yes" = "green", "no" = "red")) + theme_minimal() # this boxplot shows  
that there is a wider distribution of women representation in autocracies compared to democracies. There  
are also 3 more outliers for democracies than autocracies.
```