## Intro R and workshop 1

## Fall 2022

# Setup -------------------------------

# Load packages

library(tidyverse)

# Forgot to install?

# Run install.packages('tidyverse') & try again

# Set working directory

setwd("~/Documents/SIS 600/tesler workshop")

# adjust the path to match your project folder

# to check your current directory: `getwd()`

# use `list.files()` to see if your dataset is there

# Import 2012 ANES survey data

load('anes2012 workshop.RData')

# Use `load()` wtih .rdata.

# Use, e.g., `read.csv()` with .csv files. Etc

# Summary stats -----------------------

# Describe attitudes about labor unions (ft.unions)

summary(df$age)

# syntax: summary(object.name$variable.name)

# try it for the whole frame: summary(df)

df %>%

summarise(

Avg = mean(ft.unions, na.rm = T), # always add `na.rm=T`

SD = sd(ft.unions, na.rm = T)

)

# a customizable approach using the pipe `%>%`

# Visualize the dist

hist(df$age, xlab = 'Age (yrs)', main = 'Respondent age, 2012 ANES')

boxplot(df$age, horizontal = TRUE)

# Counts/Freq -------------------------

# Describe attitudes about global warming

tab1 = # create freq table

df %>%

count(global.warming) %>%

mutate(Percent = 100 \* n/sum(n))

# last line calculates n as relative freq

tab1

# visualize

barplot(Percent ~ global.warming, tab1)

# syntax: barplot(bar.height ~ group, object.name)

# Relationships -----------------------

# Scatter age over resentment

plot(resent.index ~ age, df)

# syntax: plot(outcome.var ~ exposure.var, object.name)

plot(jitter(resent.index) ~ jitter(age), df)

# notice the `jitter` is meant to spread overlapping points

# Estimate linear association

est1 = lm(resent.index ~ age, df)

# syntax: lm(outcome.var ~ exposure.var, object.name)

est1 # view the regression estimates

# View the plot with the regression line

plot(jitter(resent.index) ~ jitter(age), df)

abline(est1, col = 'red', lw = 3)

# `abline()` ads the line stored in est1

# Is that a strong relationship? Consider this:

plot(resent.index ~ age, df, col = 'white')

abline(h = mean(df$resent.index), col = 'blue', lwd=4)

abline(est1, col = 'red', lwd = 4, lt=2)