

HW6TEST_Jeffrey

Jack Jeffrey

Homework 6

2.1 Describe Y: PR system

```
setwd("/Users/jackjeffrey/Documents/Poli502_Jeffrey/Data")
world.data <- read.csv("world.csv")
# load data
table(world.data$pr_sys)
```

```
  No  Yes
124   67
```

```
# view table
```

```
# change labels
world.data$pr <- factor(world.data$pr_sys,
                        levels = c("Yes", "No"),
                        labels = c("PR System", "Non-PR"))
table(world.data$pr)
```

```
PR System    Non-PR
      67         124
```

```
# view table for variables
```

2.2 Describe X: Ethnic fractionalization

```
table(world.data$frac_eth3)
```

| | High | Low | Medium |
|---|------|-----|--------|
| 3 | 62 | 62 | 64 |

```
# view table  
tail(world.data[c("country", "frac_eth3")])
```

| | country | frac_eth3 |
|-----|---------------------|-----------|
| 186 | Vietnam | Low |
| 187 | Western Samoa | Low |
| 188 | Yemen | |
| 189 | Serbia & Montenegro | Medium |
| 190 | Zambia | High |
| 191 | Zimbabwe | Medium |

```
# view table with country names  
world.data$frac_eth3_ord <- factor(world.data$frac_eth3,  
                                   levels = c("Low", "Medium", "High"), # order it in this way  
                                   ordered = TRUE)  
table(world.data$frac_eth3_ord)
```

| Low | Medium | High |
|-----|--------|------|
| 62 | 64 | 62 |

```
# order the variable names  
table(world.data$frac_eth3)
```

| | High | Low | Medium |
|---|------|-----|--------|
| 3 | 62 | 62 | 64 |

```
table(world.data$frac_eth3_ord)
```

```

Low Medium High
62      64      62

```

```

# compare tables
table(world.data$frac_eth3, world.data$frac_eth3_ord)

```

```

      Low Medium High
High    0      0    62
Low     62      0     0
Medium  0      64     0

```

```
# created two-by-two table
```

2.3 Describe X-Y

```

# Ensure gmodels is installed and loaded
if (!require("gmodels")) install.packages("gmodels")

```

Loading required package: gmodels

```

library(gmodels)

# Create CrossTable for PR system and ethnic fractionalization
CrossTable(world.data$pr, world.data$frac_eth3_ord)

```

```

      Cell Contents
|-----|
|              N |
| Chi-square contribution |
|      N / Row Total |

```

| | |
|-------|-----------------|
| | N / Col Total |
| | N / Table Total |
| ----- | |

Total Observations in Table: 188

| | world.data\$frac_eth3_ord | | | |
|----------------|---------------------------|--------|-------|-----------|
| world.data\$pr | Low | Medium | High | Row Total |
| ----- | | | | |
| PR System | 23 | 27 | 15 | 65 |
| | 0.114 | 1.073 | 1.932 | |
| | 0.354 | 0.415 | 0.231 | 0.346 |
| | 0.371 | 0.422 | 0.242 | |
| | 0.122 | 0.144 | 0.080 | |
| ----- | | | | |
| Non-PR | 39 | 37 | 47 | 123 |
| | 0.060 | 0.567 | 1.021 | |
| | 0.317 | 0.301 | 0.382 | 0.654 |
| | 0.629 | 0.578 | 0.758 | |
| | 0.207 | 0.197 | 0.250 | |
| ----- | | | | |
| Column Total | 62 | 64 | 62 | 188 |
| | 0.330 | 0.340 | 0.330 | |
| ----- | | | | |

```
# Optional: Run without additional parameters to isolate the issue
CrossTable(world.data$pr, world.data$frac_eth3_ord,
  prop.r = FALSE,
  prop.t = FALSE,
  prop.chisq = FALSE,
  chisq = TRUE) # Perform chi-square test
```

| | |
|---------------|---------------|
| Cell Contents | |
| ----- | |
| | N |
| | N / Col Total |

|-----|

Total Observations in Table: 188

| | world.data\$frac_eth3_ord | | | |
|----------------|---------------------------|-------------|-------------|-----------|
| world.data\$pr | Low | Medium | High | Row Total |
| PR System | 23 0.371 | 27 0.422 | 15 0.242 | 65 |
| Non-PR | 39 0.629 | 37 0.578 | 47 0.758 | 123 |
| Column Total | 62 0.330 | 64 0.340 | 62 0.330 | 188 |

Statistics for All Table Factors

Pearson's Chi-squared test

Chi^2 = 4.767842 d.f. = 2 p = 0.09218841

Interpretation: chi-squared equals 4.76 with 2 degrees of freedom and a p-value of
0.092. We could reject the null at a 90% confidence level.

3. Correlation

3.1 Describe Y: labor freedom

```
summary(world.data$free_labor)
```

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. | NA's |
|-------|---------|--------|-------|---------|-------|------|
| 20.00 | 50.10 | 60.80 | 62.08 | 75.90 | 98.90 | 18 |

```
# view statistics for free labor variable, variable has 18 NA's
if (!require("tidyverse")) install.packages("tidyverse")
```

Loading required package: tidyverse

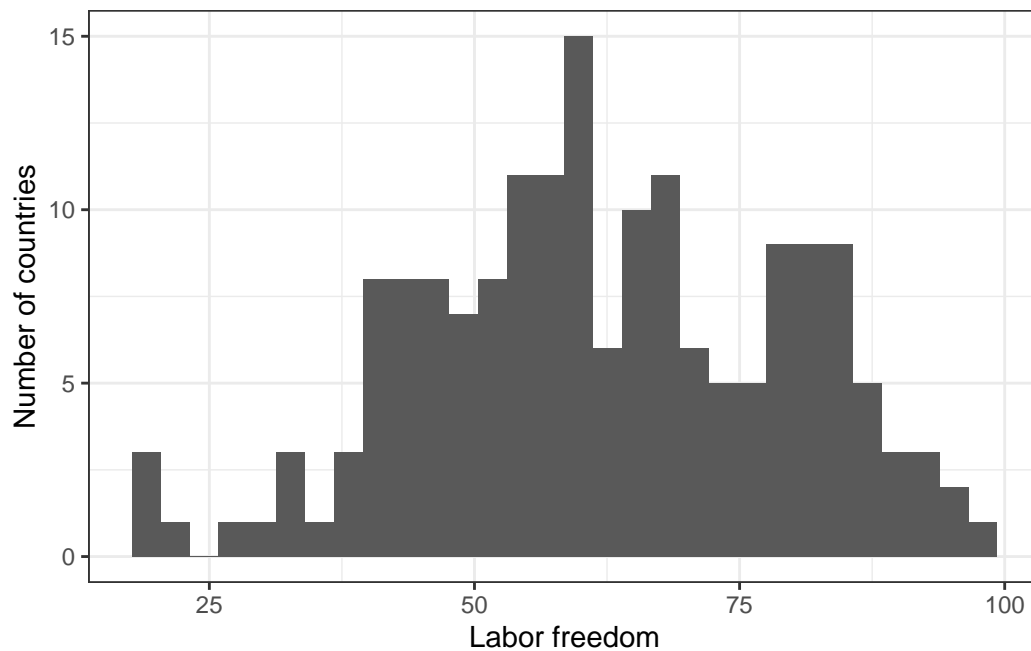
```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.4      v readr      2.1.5
v forcats    1.0.0      v stringr    1.5.1
v ggplot2     3.5.1      v tibble     3.2.1
v lubridate  1.9.3      v tidyr      1.3.1
v purrr       1.0.2
```

```
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

```
library(tidyverse)
# loaded tidyverse package
g <- ggplot(world.data, aes(x = free_labor)) + geom_histogram()
g <- g + ylab("Number of countries") + xlab("Labor freedom") + theme_bw()
g
```

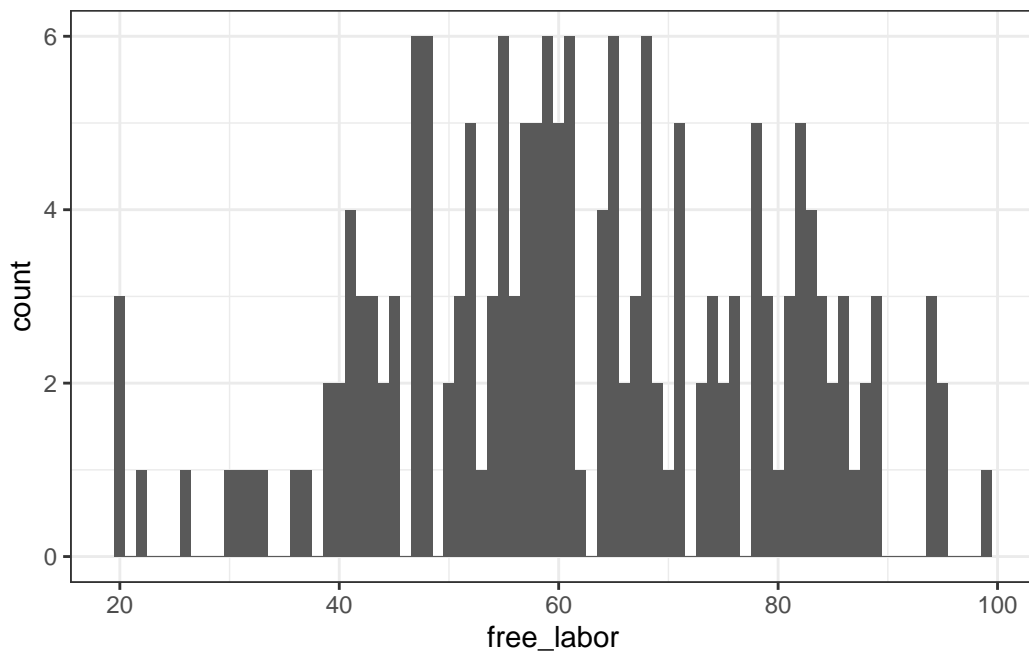
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

Warning: Removed 18 rows containing non-finite outside the scale range
(`stat_bin()`).



```
# created a graphical summary of number of countries and labor freedom
cleaned_data <- world.data[is.finite(world.data$free_labor), ]
# removed NAs in new object
cleaned_data <- ggplot(world.data, aes(x = free_labor)) +
  geom_histogram(binwidth = 1) +
  theme_bw()
cleaned_data
```

Warning: Removed 18 rows containing non-finite outside the scale range (`stat_bin()`).



```
# graphed variable without NAs and new binwidth value
```

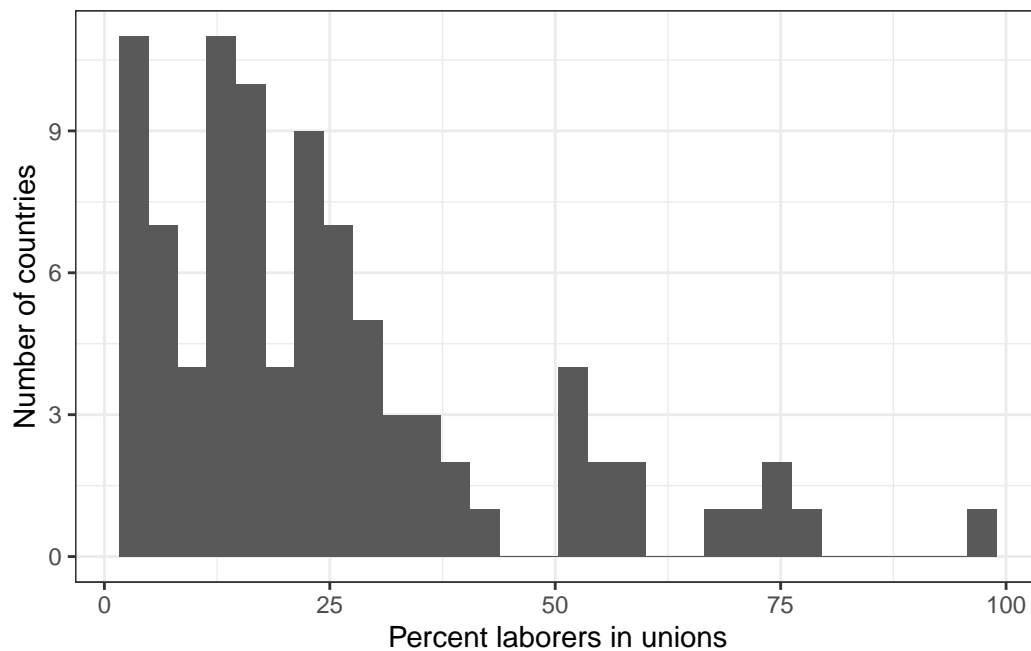
3.2 Describe X: union density

```
summary(world.data$unions)
```

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. | NA's |
|------|---------|--------|-------|---------|-------|------|
| 2.00 | 11.45 | 19.10 | 24.74 | 30.80 | 96.10 | 100 |

```
# numerical summary for union variable, 100 NAs found
g <- ggplot(world.data, aes(x = unions)) +
  geom_histogram(na.rm = TRUE) + # use statbin to remove NAs
  ylab("Number of countries") +
  xlab("Percent laborers in unions") +
  theme_bw()
g
```

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

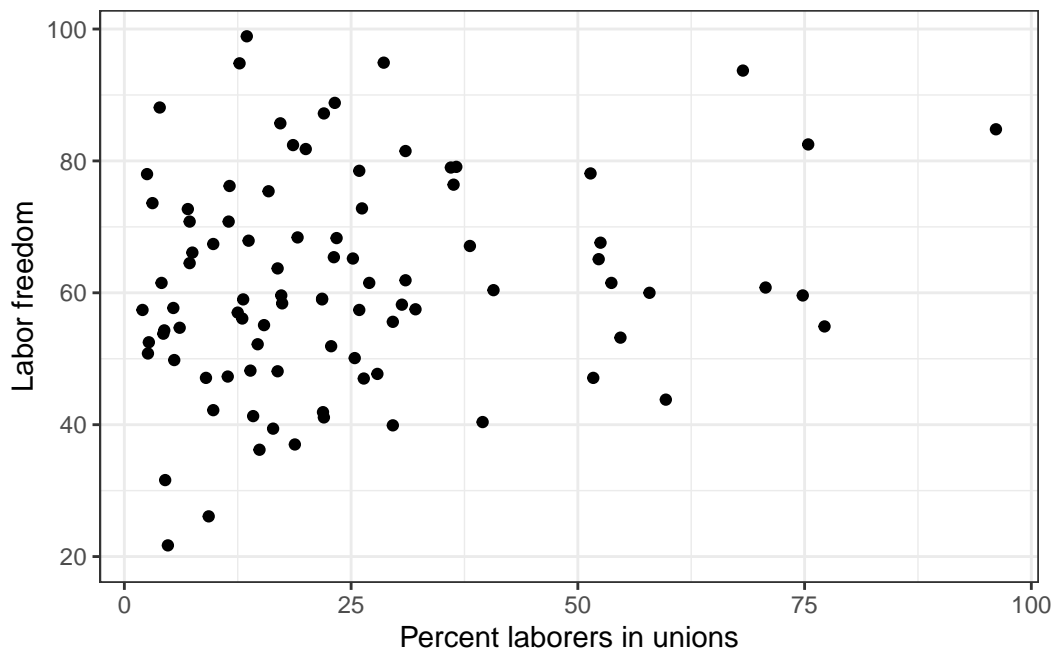


```
# saved the plot
```

3.3 Describe X-Y

```
g <- ggplot(world.data, aes(x = unions, y = free_labor)) + geom_point() + theme_bw()
g <- g + ylab("Labor freedom") + xlab("Percent laborers in unions")
g
```

Warning: Removed 100 rows containing missing values or values outside the scale range (`geom_point()`).



```
# plotted an X-Y scatterplot graph using geom_point
```

Calculate test-statistic & p

```
lab.uni.mat <- as.matrix(world.data[c("unions", "free_labor")])
# create a new matrix
if (!require("Hmisc")) install.packages("Hmisc")
```

Loading required package: Hmisc

Attaching package: 'Hmisc'

The following objects are masked from 'package:dplyr':

src, summarize

The following objects are masked from 'package:base':

format.pval, units

```
library(Hmisc)
# loaded Hmisc package
rcorr(lab.uni.mat, type = "pearson")
```

```

      unions free_labor
unions      1.00      0.18
free_labor  0.18      1.00
```

```

n
      unions free_labor
unions      91      91
free_labor  91     173
```

```

P
      unions free_labor
unions      0.0913
free_labor 0.0913
```

```
# Interpretation: for the Pearson's test we see that the correlation coefficient is
# positive at 0.18, with a total of 91 observations present in both X and Y variables.
# The p-value of 0.0913 indicates that our expectations of more labor unions correlating
# to more labor freedom is statistically significant at the 90% level but not at the 95%.
```

4. Adding lines and labels to a graph

```
lab.data <- world.data[!is.na(world.data$unions) & !is.na(world.data$free_labor), ]
# created a new object with NAs removed
mean.y <- mean(lab.data$free_labor)
mean.x <- mean(lab.data$unions)

mean.y
```

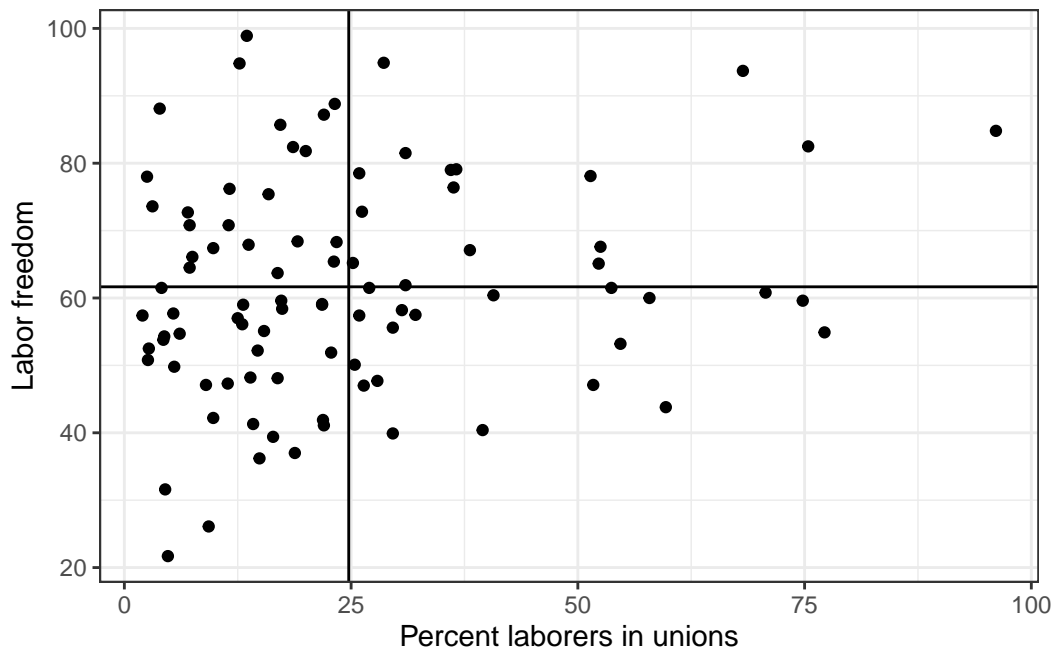
```
[1] 61.65714
```

```
mean.x
```

```
[1] 24.73846
```

```
# calculate the means - mean for Y equals 61.66, mean for X equals 24.74

g <- ggplot(lab.data, aes(x = unions, y = free_labor)) + geom_point() + theme_bw()
g <- g + ylab("Labor freedom") + xlab("Percent laborers in unions")
g <- g + geom_hline(aes(yintercept = mean.y))
g <- g + geom_vline(aes(xintercept = mean.x))
g
```



```
# scatterplot using hline and vline to draw lines at the X and Y means

# Optional: Add country labels for readability using ggrepel
if (!require("ggrepel")) install.packages("ggrepel")
```

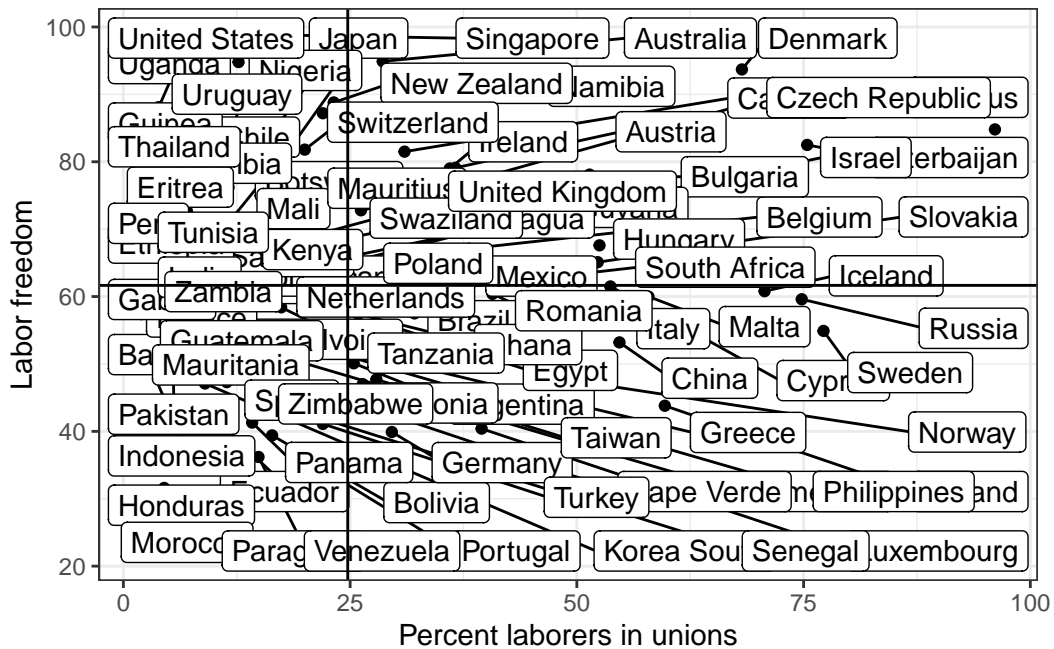
Loading required package: ggrepel

```
library(ggrepel)
options(ggrepel.max.overlaps = Inf)

g <- ggplot(lab.data, aes(x = unions, y = free_labor)) +
  geom_point() +
  geom_label_repel(aes(label = country)) +
  ylab("Labor freedom") + xlab("Percent laborers in unions") +
```

```
geom_hline(aes(yintercept = mean.y)) +
geom_vline(aes(xintercept = mean.x)) +
theme_bw()
```

g



Final scatterplot with country labels for readability