First lab:

email50\_updated <- email50 |>

mutate(num\_char\_cut=if\_else(

email50\_updated <- email50 |>

mutate(num\_char\_cut=if\_else(

num\_char >= med,

email50\_updated <- email50 |>

mutate(num\_char\_cut=if\_else(

num\_char >= med,

'at or above median',

email50\_updated <- email50 |>

mutate(num\_char\_cut=if\_else(

num\_char >= med,

'at or above median',

'below median'))

summary(email50\_updated$num\_char\_cut)

---------------------------------------------------------------------------------------------------------------------

US Government

Los Angeles Crimes 2020-2023

install.packages('openintro')

install.packages('tidyverse')

my\_data\_set <- read.csv("Crime\_Data\_from\_2020\_to\_Present.csv")

library(openintro)

library(tidyverse)

library(dplyr)

head(my\_data\_set)

glimpse(my\_data\_set)

my\_data\_set\_filtered <- my\_data\_set %>%

filter(AREA > 10)

glimpse(my\_data\_set\_filtered)

dim(my\_data\_set\_filtered)

(med <- median(my\_data\_set$AREA))

my\_data\_set\_filtered\_updated <- my\_data\_set |>

mutate(AREA\_cut=if\_else(

AREA >= med,

'at or above median',

'below median'))

my\_data\_set\_filtered\_updated$AREA\_cut

my\_data\_set\_filtered\_updated$AREA\_cut <- as.factor(my\_data\_set\_filtered\_updated$AREA\_cut)

summary(my\_data\_set\_filtered\_updated$AREA\_cut)

Second lab:

install.packages("gridExtra")

comics <- read.csv("D:/R labs/comic\_characters.csv")

library(dplyr)

library(tidyr)

library(ggplot2)

library(gridExtra)

head(comics)

levels(comics$align)

levels(comics$sex)

contingency\_table <- comics %>%

count(sex, align) %>%

pivot\_wider(names\_from = align, values\_from = n)

contingency\_table

class(comics$align)

comics$align <- as.factor(comics$align)

comics\_filtered <- comics %>%

filter(align != "Lowest\_Level")

comics\_filtered$align <- droplevels(comics\_filtered$align)

levels(comics\_filtered$align)

plot1 <- ggplot(comics, aes(x = align, fill = sex)) +

geom\_bar(position = "dodge") +

labs(x = "Alignment", y = "Count", fill = "Gender") +

ggtitle("Side-by-Side Barchart of Alignment and Gender")

plot2 <- ggplot(comics, aes(x = sex, fill = align)) +

geom\_bar(position = "dodge") +

labs(x = "Gender", y = "Count", fill = "Alignment") +

ggtitle("Side-by-Side Barchart of Gender and Alignment")

grid.arrange(plot1, plot2, ncol = 2)

---------------------------------------------------------------------------------------------------------------------

install.packages("openintro")

install.packages("ggplot2")

library(openintro)

library(ggplot2)

data(cars93)

colnames(cars93)

plot1 <- ggplot(cars93, aes(x = price)) +

geom\_histogram(binwidth = 3, fill = "blue", color = "black") +

labs(title = "Histogram of Horsepower (Binwidth = 3)",

x = "Horsepower",

y = "Frequency")

plot2 <- ggplot(cars93, aes(x = price)) +

geom\_histogram(binwidth = 30, fill = "green", color = "black") +

labs(title = "Histogram of Horsepower (Binwidth = 30)",

x = "Horsepower",

y = "Frequency")

plot3 <- ggplot(cars93, aes(x = price)) +

geom\_histogram(binwidth = 60, fill = "red", color = "black") +

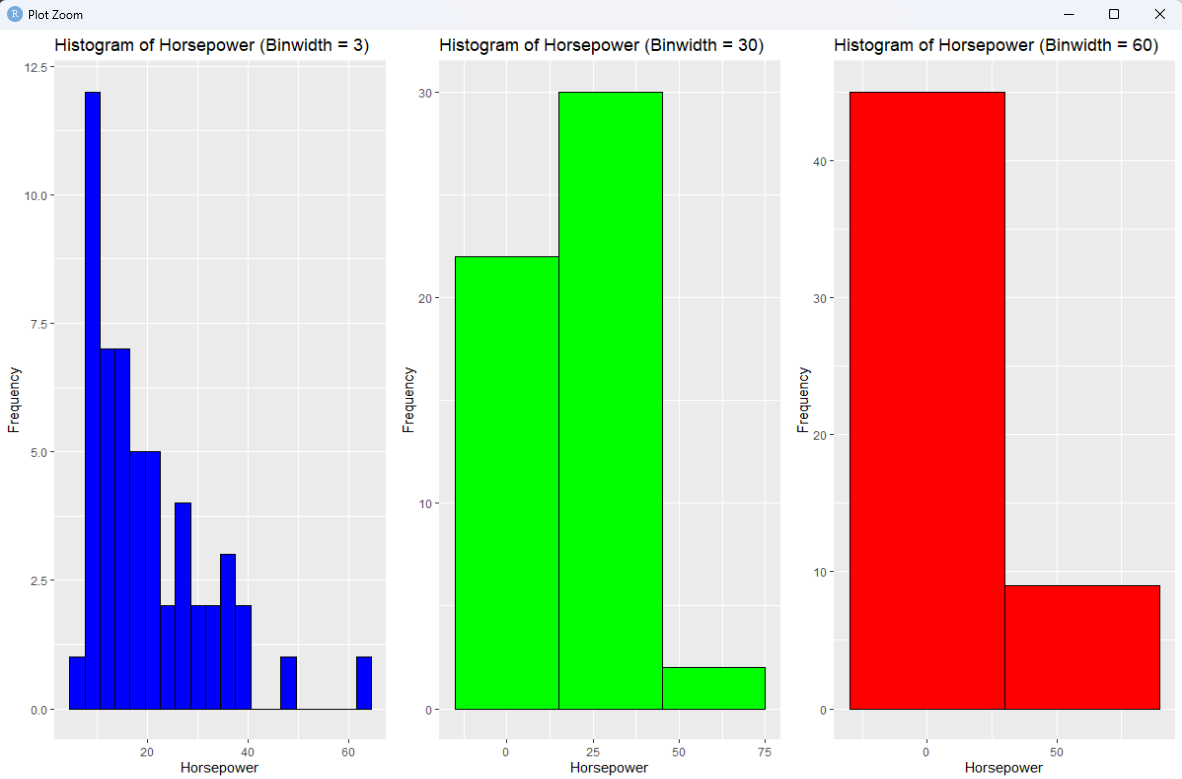
labs(title = "Histogram of Horsepower (Binwidth = 60)",

x = "Horsepower",

y = "Frequency")

library(gridExtra)

grid.arrange(plot1, plot2, plot3, ncol = 3)



library(openintro)

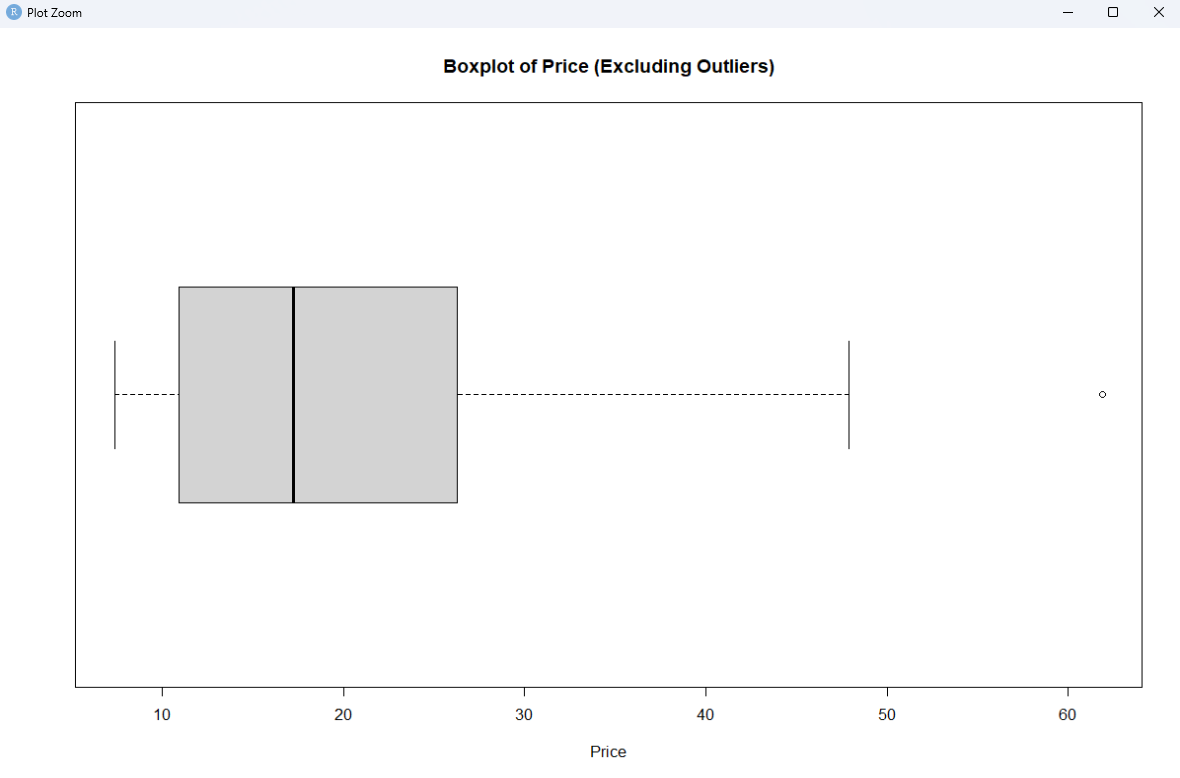
library(ggplot2)

data(cars93)

boxplot(cars93$price, horizontal = TRUE, main = "Boxplot of Price", xlab = "Price")

cars\_no\_out <- cars93[cars93$price < 40000, ]

boxplot(cars\_no\_out$price, horizontal = TRUE, main = "Boxplot of Price (Excluding Outliers)", xlab = "Price")



library(openintro)

library(ggplot2)

data(cars93)

ggplot(cars93, aes(x = mpg\_city)) +

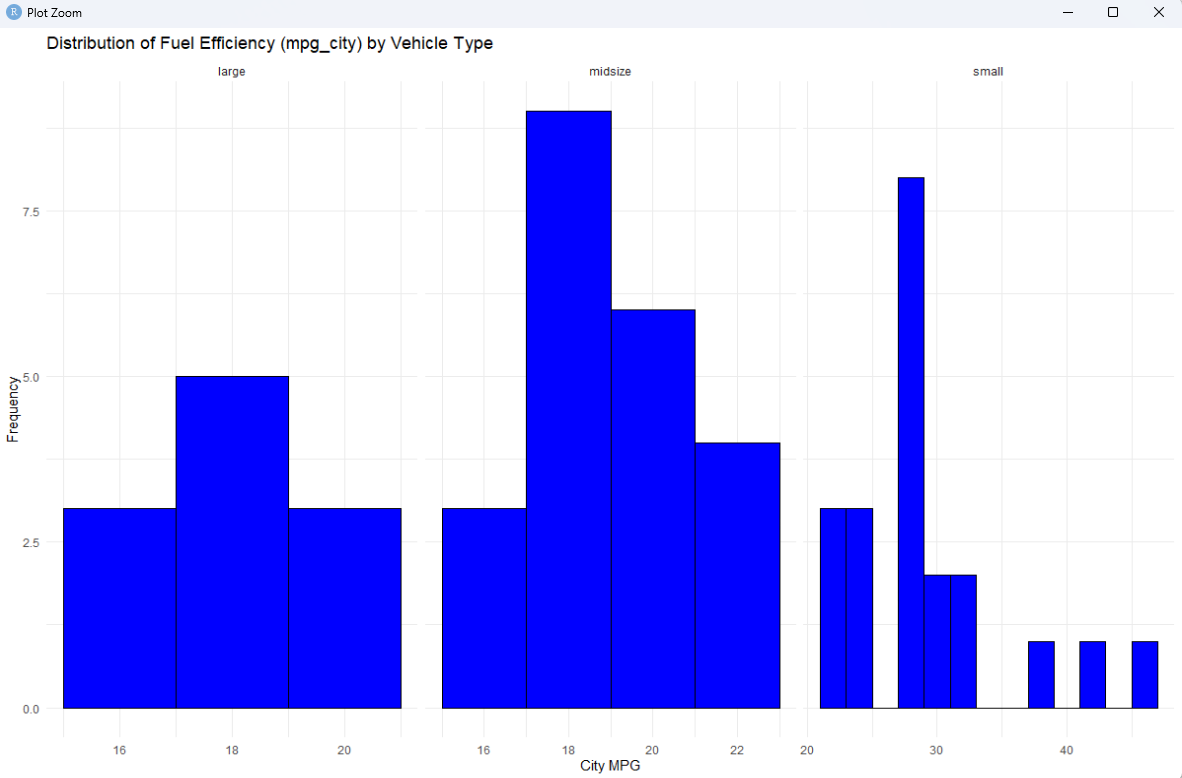
geom\_histogram(binwidth = 2, fill = "blue", color = "black") +

facet\_wrap(~ type, scales = "free\_x") +

labs(x = "City MPG", y = "Frequency") +

ggtitle("Distribution of Fuel Efficiency (mpg\_city) by Vehicle Type") +

theme\_minimal()



library(openintro)

library(ggplot2)

data(cars93)

cars93 %>%

# Step 2: Pipe the data into the filter() function

filter(price < 30000) %>%

# Step 4: Pipe the filtered data into the ggplot() function

ggplot(aes(x = mpg\_city)) +

# Step 5: Specify the variables and wrap the histogram by drive\_train variable

geom\_histogram() +

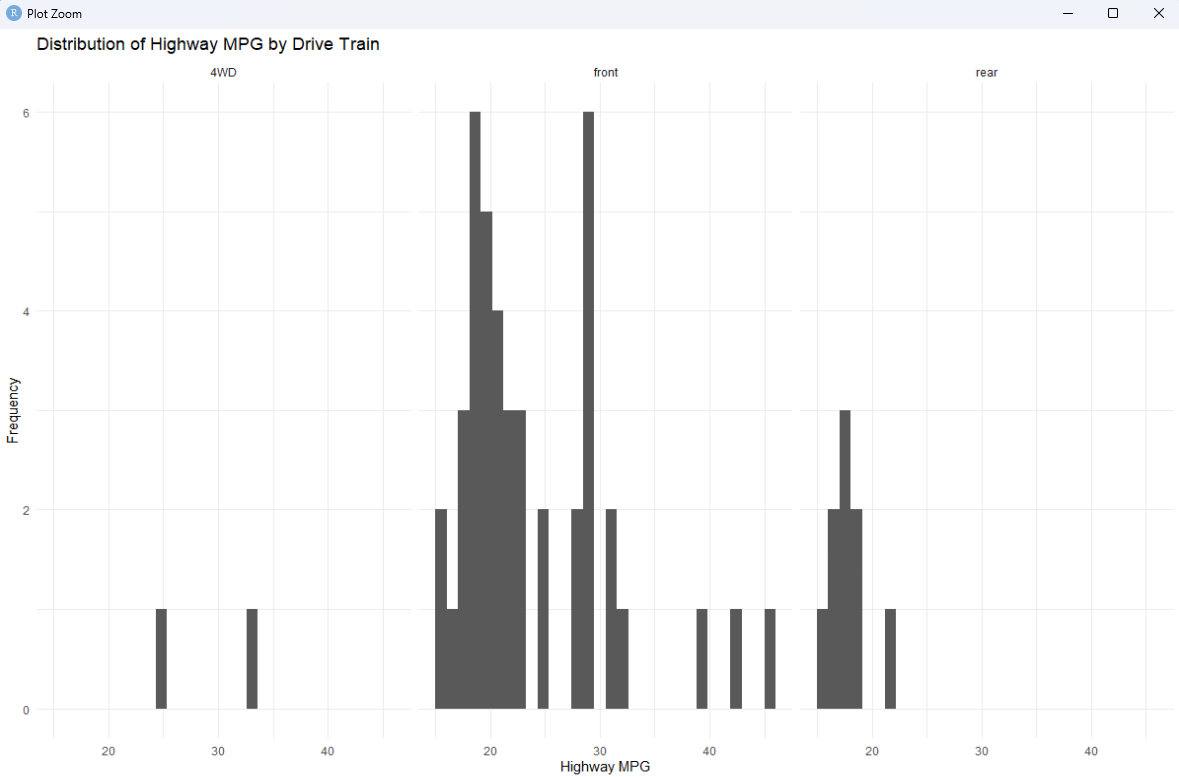
facet\_wrap(~ drive\_train) +

# Step 6: Add a histogram layer to the plot

labs(x = "Highway MPG", y = "Frequency") +

ggtitle("Distribution of Highway MPG by Drive Train") +

theme\_minimal()



Third lab:

install.packages("gapminder")

library(gapminder)

library(ggplot2)

gap2007. <- gapminder %>%

filter(year == 2007)

life\_expectancy\_summary <- gap2007 %>%

group\_by(continent) %>%

summarize(mean\_life\_expectancy = mean(lifeExp),

median\_life\_expectancy = median(lifeExp))

ggplot(gap2007, aes(x = continent, y = lifeExp)) +

geom\_boxplot() +

labs(x = "Continent", y = "Life Expectancy") +

ggtitle("Life Expectancy by Continent in 2007")

spread\_summary <- gap2007 %>%

group\_by(continent) %>%

summarize(iqr\_life\_expectancy = IQR(lifeExp))

spread\_summary <- gap2007 %>%

group\_by(continent) %>%

summarize(sd\_life\_expectancy = sd(lifeExp),

iqr\_life\_expectancy = IQR(lifeExp),

num\_countries = n())

ggplot(gap2007, aes(x = lifeExp, fill = continent)) +

geom\_density(alpha = 0.3) +

labs(x = "Lifespan", y = "Density") +

ggtitle("Life expectancy density by continent in 2007")

americas\_data <- gap2007 %>%

filter(continent == "Americas")

center\_and\_spread <- americas\_data %>%

summarize(median\_life\_expectancy = median(lifeExp),

iqr\_life\_expectancy = IQR(lifeExp),

sd\_life\_expectancy = sd(lifeExp))

americas\_data <- americas\_data %>%

mutate(log\_population = log(pop))

head(americas\_data)

ggplot(gap2007, aes(x = pop)) +

geom\_density() +

labs(x = "Population", y = "Density") +

ggtitle("Population density in 2007")

gap2007 <- gap2007 %>%

mutate(log\_pop = log(pop))

ggplot(gap2007, aes(x = log\_pop)) +

geom\_density() +

labs(x = "log(Population)", y = "Density") +

ggtitle("Density of natural logarithm of population in 2007")

asia\_data <- gap2007 %>%

filter(continent == "Asia")

asia\_data <- asia\_data %>%

mutate(is\_outlier = if\_else(lifeExp < 50, TRUE, FALSE))

ggplot(asia\_data, aes(x = "", y = lifeExp)) +

geom\_boxplot() +

labs(x = "", y = "Lifespan") +

ggtitle("Boxplot life expectancy in Asia")

