21BDS0340

Abhinav Dinesh Srivatsa

Exploratory Data Analysis Lab

Experiment – VIII

Code:

```
library(dplyr)
library(ggplot2)
library(corrplot)

# performing exploratory data analysis with mtcars
data = mtcars
```

Output:

```
> library(dplyr)
> library(ggplot2)
> library(corrplot)
>
> # performing exploratory data analysis with mtcars
> data = mtcars
```

Code:

```
# viewing data structure and dimensions
str(data)
dim(data)
```

Output:

```
> # viewing data structure and dimensions
> str(data)
'data.frame':   32 obs. of  11 variables:
 $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
 $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
 $ disp: num  160 160 108 258 360 ...
 $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
 $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
 $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
 $ qsec: num  16.5 17 18.6 19.4 17 ...
 $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
 $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
 $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
 $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
> dim(data)
[1] 32 11
```

Code:

```
# summarising the data
summary(data)
```

Output:

```
> # summarising the data
> summary(data)
      mpg             cyl             disp             hp             drat
 Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0   Min.   :2.760
 1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5   1st Qu.:3.080
 Median :19.20   Median :6.000   Median :196.3   Median :123.0   Median :3.695
 Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7   Mean   :3.597
 3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0   3rd Qu.:3.920
 Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0   Max.   :4.930
       wt             qsec             vs              am              gear
 Min.   :1.513   Min.   :14.50   Min.   :0.0000   Min.   :0.0000   Min.   :3.000
 1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:3.000
 Median :3.325   Median :17.71   Median :0.0000   Median :0.0000   Median :4.000
 Mean   :3.217   Mean   :17.85   Mean   :0.4375   Mean   :0.4062   Mean   :3.688
 3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:4.000
 Max.   :5.424   Max.   :22.90   Max.   :1.0000   Max.   :1.0000   Max.   :5.000
      carb
 Min.   :1.000
 1st Qu.:2.000
 Median :2.000
 Mean   :2.812
 3rd Qu.:4.000
 Max.   :8.000
```

Code:

```
# missing data checking
sum(is.na(data))
```

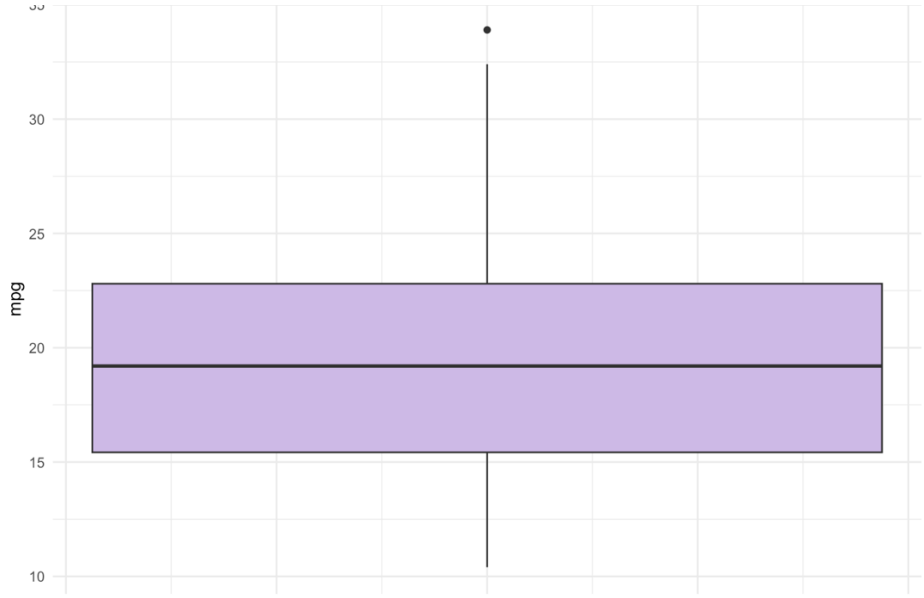
Output:

```
> # missing data checking
> sum(is.na(data))
[1] 0
```

Code:

```
# outlier detection
ggplot(data, aes(y=mpg)) +
  geom_boxplot(fill="#cebae6") +
  theme_minimal()
```

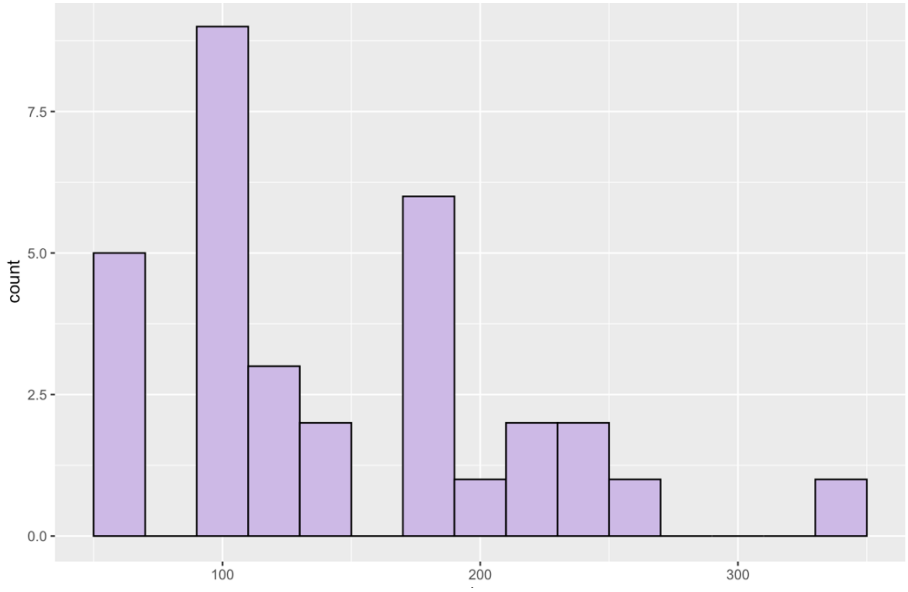
Output:



Code:

```
# univariate analysis
ggplot(data, aes(x=hp)) +
  geom_histogram(binwidth=20, fill="#cebae6", color="black")
```

Output:

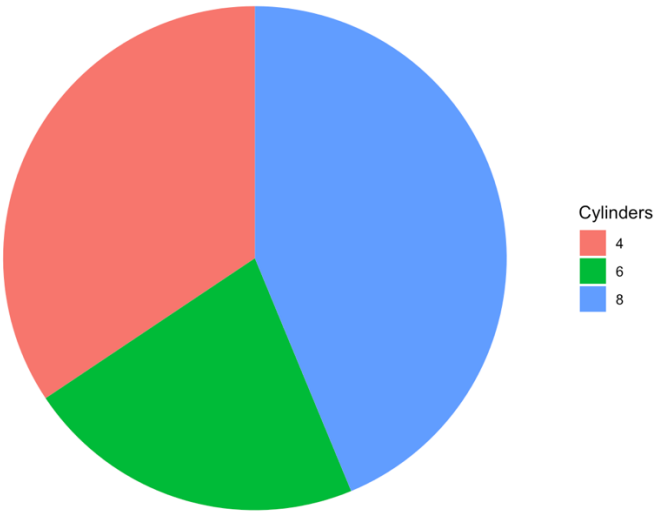


Code:

```
# univariate analysis
cyl_counts <- as.data.frame(table(data$cyl))
colnames(cyl_counts) <- c("Cylinders", "Count")
ggplot(cyl_counts, aes(x="", y=Count, fill=Cylinders)) +
  geom_bar(stat="identity", width=1) +
  coord_polar(theta = "y") +
```

```
  theme_void()
```

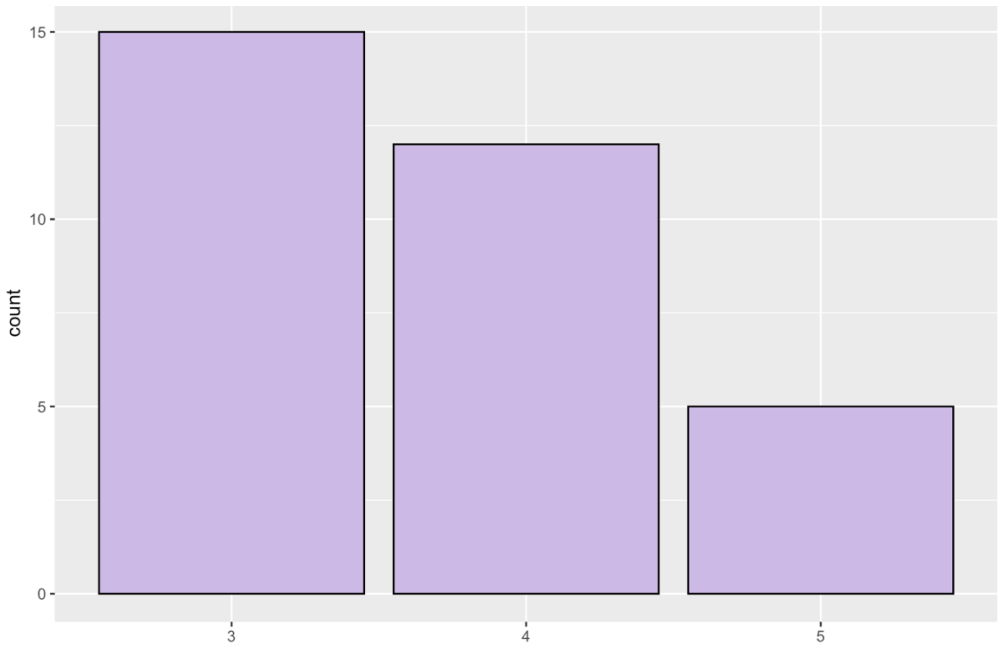
Output:



Code:

```
data$gear = as.factor(data$gear)
ggplot(data, aes(x=gear)) +
  geom_bar(fill="#cebae6", color="black")
```

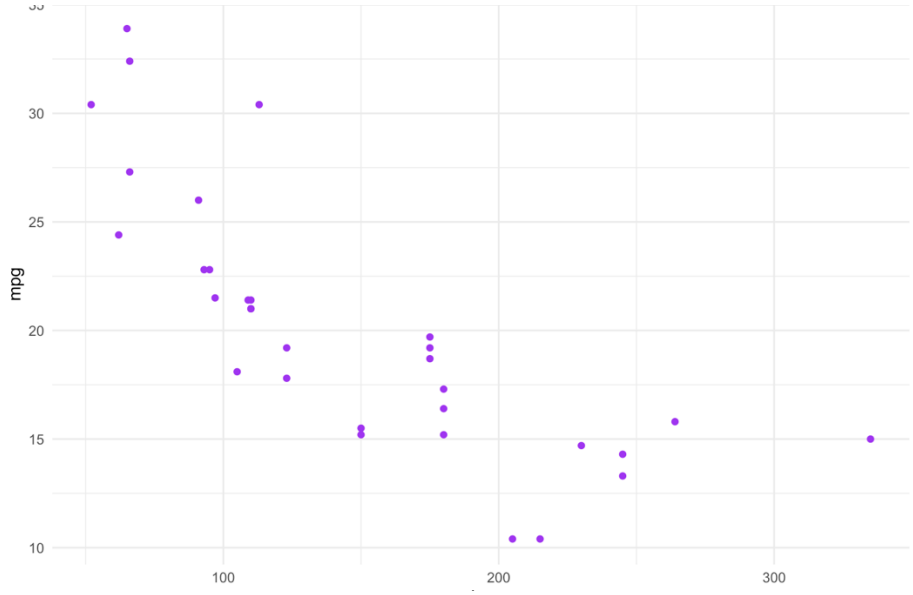
Output:



Code:

```
# bivariate analysis
ggplot(data, aes(x=hp, y=mpg)) +
  geom_point(color="purple") +
  theme_minimal()
```

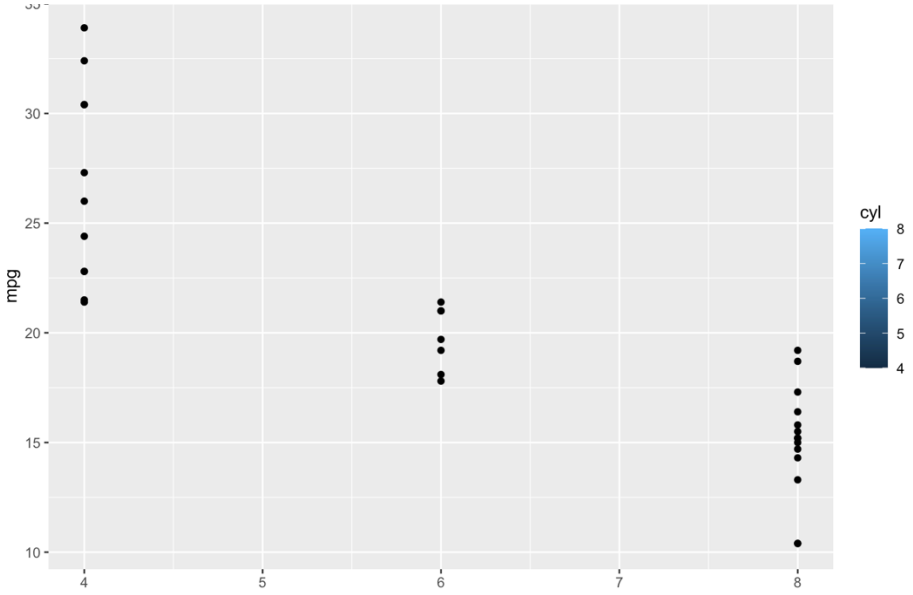
Output:



Code:

```
ggplot(data, aes(x=cyl, y=mpg, fill=cyl)) +
  geom_point()
```

Output:



Code:

```
# multivariate analysis
data$cyl = as.factor(data$cyl)
data$gear = as.factor(data$gear)
```

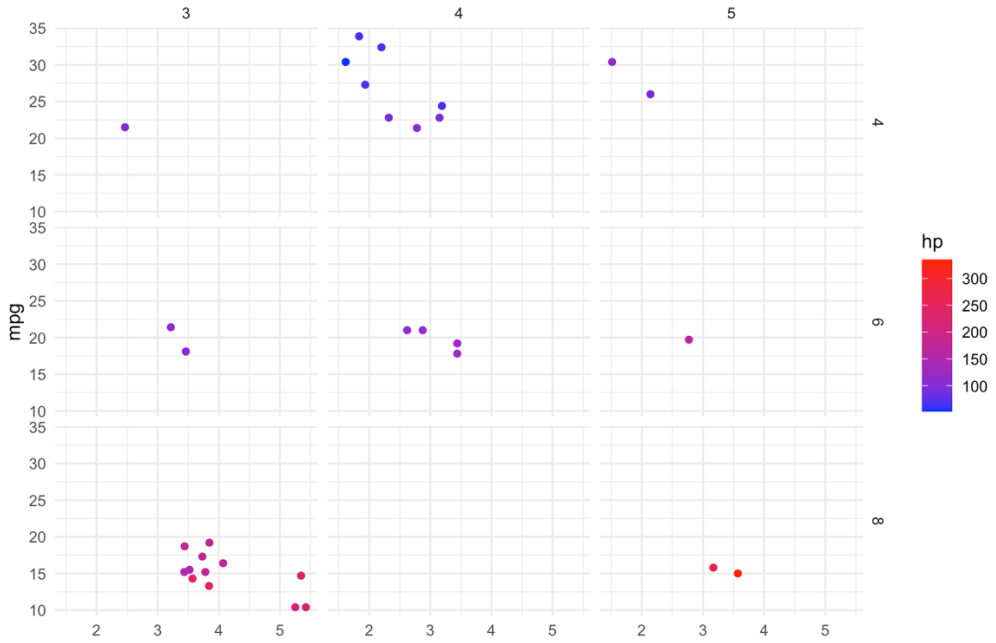
Output:

```
> # multivariate analysis
> data$cyl = as.factor(data$cyl)
> data$gear = as.factor(data$gear)
```

Code:

```
ggplot(data, aes(x=wt, y=mpg, color=hp)) +
  geom_point() +
  scale_color_gradient(low="blue", high="red") +
  facet_grid(cyl~gear) +
  theme_minimal()
```

Output:



Code:

```
# correlation analysis
cor_matrix = cor(data %>% select_if(is.numeric))
corrplot(cor_matrix)
```

Output: