PSTAT131 - main - RStudio

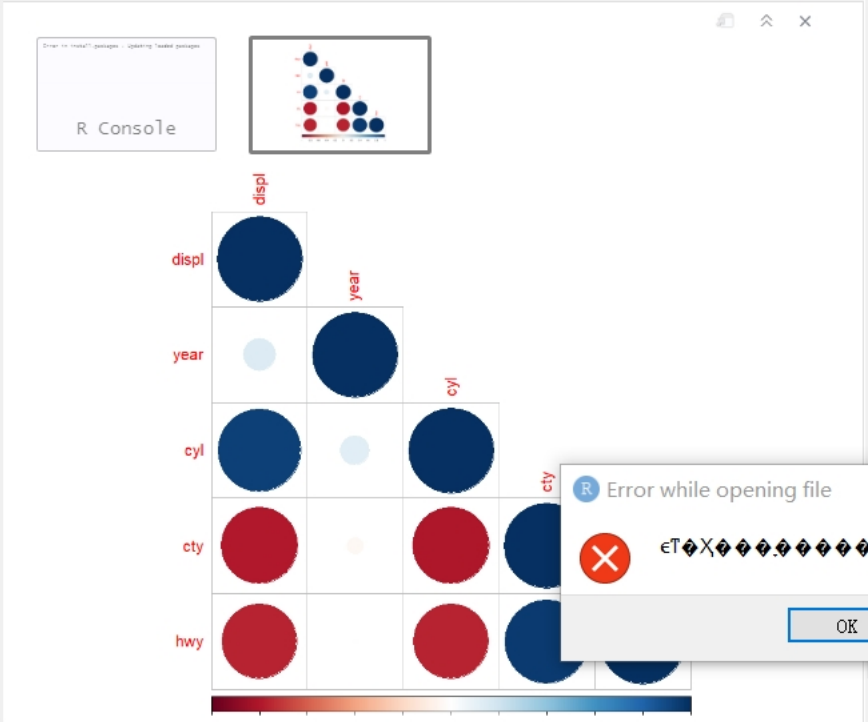
File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | PSTAT131

131 HW1.Rmd | 1.R | 126 HW4.Rmd

Knit on Save | Knit | Run | Source | Visual | Outline

```
fl, class))
cormpg=cor(newmpg)
corrplot(cormpg,type="lower")
```



Error while opening file

OK

68:1 | Chunk 2 | R Markdown

Environment | History | Connections | Git | Tutorial

Diff | Commit | main

Staged	Status	Path
	M	131 HW1.Rmd
	?	131-HW1.Rmd

Files | Plots | Packages | Help | Viewer

Zoom | Export

Console | Terminal | Render | Jobs

.../PSTAT131/131 HW1.Rmd | Output | Issues

C:/Users/12191/Desktop/PSTAT131/PSTAT131/131-HW1.Rmd

Line 98 contrib.url(repos, "source")
trying to use CRAN without setting a mirror
<Anonymous> ... withVisible -> eval -> eval -> install.pa

PSTAT131 - main - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | PSTAT131

131 HW1.Rmd | 1.R | 126 HW4.Rmd

Knit on Save | Knit | Run

Source | Visual | Outline

```
1  ---
2  title: "131 Homework1"
3  author: "Scott Shang (8458655)"
4  date: "`r format(Sys.Date(), '%B %d, %Y')`"
5  output:
6    html_document:
7      df_print: paged
8    pdf_document: default
9  ---
10
11 ---
12 title: "131 Homework1"
13 author: "Scott Shang (8458655)"
14 date: "`r format(Sys.Date(), '%B %d, %Y')`"
15 output: pdf_document
16 ---
17
18
19 ```{r setup, echo=FALSE}
20 library(knitr)
21 # set global chunk options: images will be 7x5 inches
22 knitr::opts_chunk$set(fig.width=7, fig.height=5)
23 options(digits = 4)
24
25
26 ## indents are for indenting r code as formatted text
27 ## They may need to be adjusted depending on your OS
28 # if your output looks odd, increase or decrease indent
29 indent1 = '    '
30 indent2 = '        '
31 indent3 = '            '
32 ```
33
34
35 Question 1:
36 Supervised learning: Machine learning that including prediction, estimation,
   model selection, and inference.
37 To be specific, supervised learning can accurately predict future response given
   predictors, understand how predictors affect response, select the best model,
   and assess the quality of predictions and estimation.
38 Unsupervised learning don't have a supervisor, and the model work on its own.
   (From Lec1 pg30)
39
40 Question 2:
41 In the context of machine learning, the regression model is where the response
   variable Y is quantitative. To be specific, Y are numerical values.
42 For the classification model, the response variable Y is qualitative, i.e.
   categorical values. (Lec1 31)
43
44 Question 3:
45 Didn't covered.
46
47 Question 4:
48 Descriptive models: Choose model to best visually emphasize a trend in data.
49 Inferential models focus on those features are essential. It is aimed to test
   theories and possibly causal claims. It also state the relationship between
   outcome and predictors.
50 Predictive models answer what combo of features fits the model best. And it aims
   to predict Y with minimum reducible error. It doesn't focus on hypothesis tests.
51 (Lec2 pg7)
```

68:1 | Chunk 2 | R Markdown

Console

Environment | History | Connections

Diff | Commit

Staged	Status	Path
☐	M	131 HW1.Rmd

Files | Plots | Packages | Help | Viev

Zoom | Export

PSTAT131 - main - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | PSTAT131

131 HW1.Rmd | 1.R | 126 HW4.Rmd

Knit on Save | Knit | Run

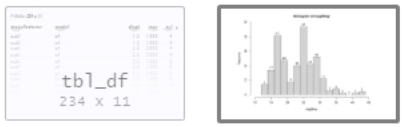
Source | Visual | Outline

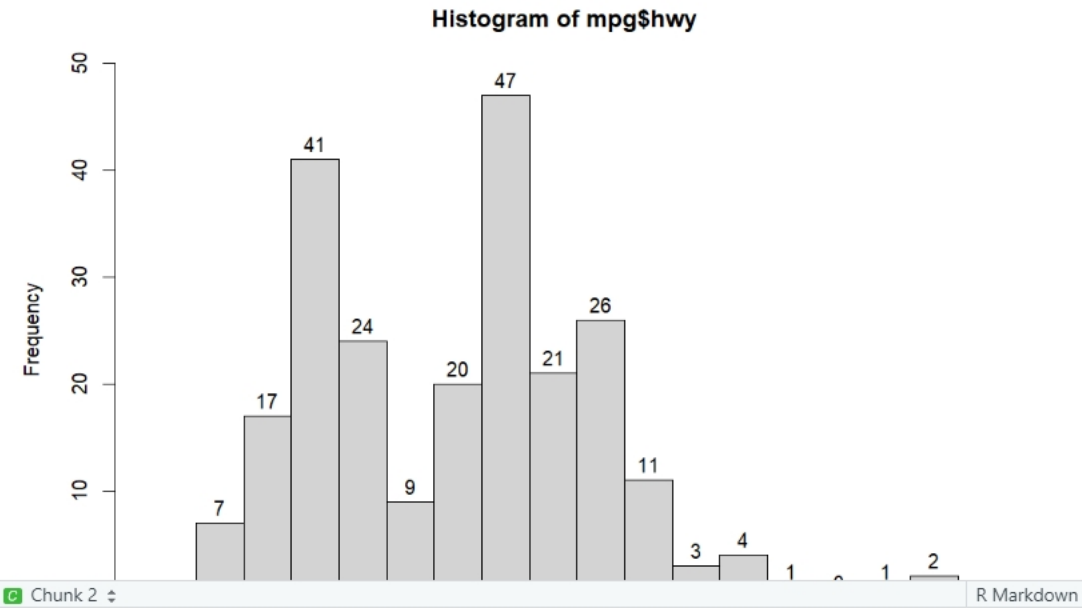
```
Predictive models answer what combo of features fits the model best. And it aims
to predict Y with minimum reducible error. It doesn't focus on hypothesis tests.
(Lec2 pg7)

Question 5:
Mechanistic assume a parametric form. It relies on known formula and parameters.
It can add parameters to add more flexibility, but too many parameters may
overfit the model. It won't match true unknown f.
Empirically-driven means the model has no assumptions about f. It require a
large number of observations. By default, it has much more flexibility than
mechanistic models. It also has the risk of overfitting.
(Lec2 pg6)
bias-variance tradeoff didn't covered in class.

Question 6:
The first question is predictive, since it using a set of information to predict
outcome(who will they vote).
The second question is inferential. It want to test whether the feature 'having
contact with the candidate' is significant regarding the outcome. We want to
observed the pattern with that feature.
(Lec2 pg7)


Ex1:
```{r}
library(tidyverse)
mpg
h=hist(mpg$hwy,xlim=c(10,45),ylim=c(0,50),breaks=20)
text(h$mids,h$counts,labels=h$counts, adj=c(0.5, -0.5))
```
```





68:1 | Chunk 2 | R Markdown

Console

Environment | History | Connections

Diff | Commit

Staged | Status | Path

131 HW1.Rmd

Files | Plots | Packages | Help | Viev

Zoom | Export

PSTAT131 - main - RStudio

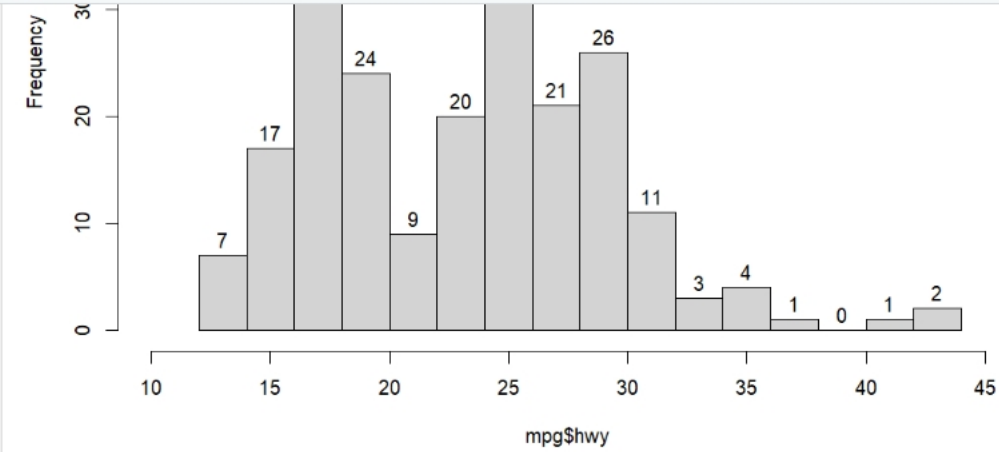
File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | PSTAT131

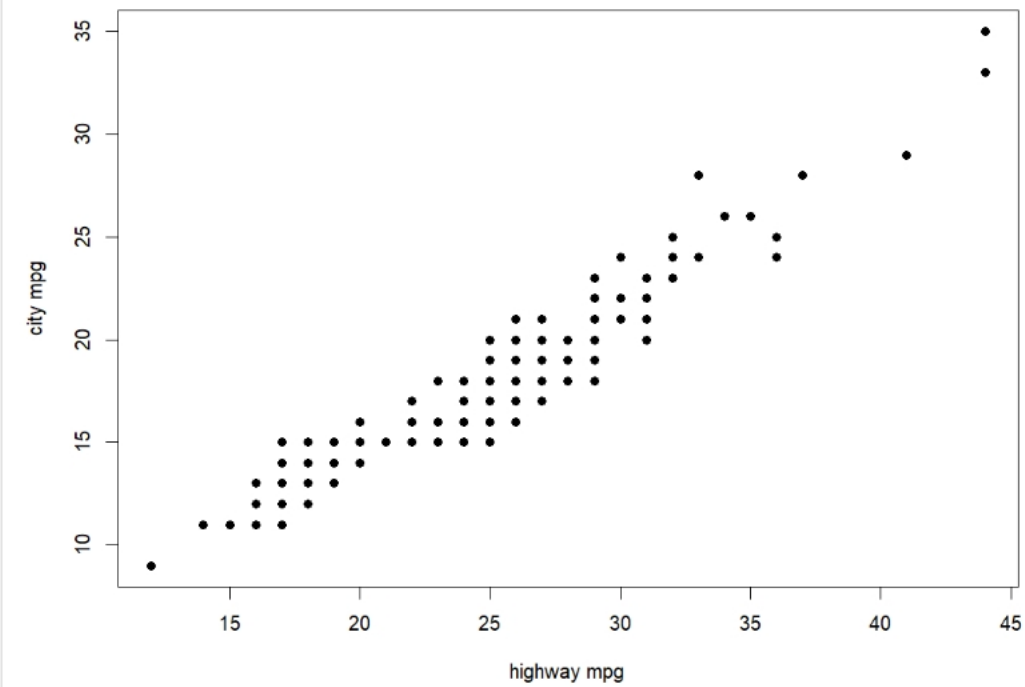
131 HW1.Rmd | 1.R | 126 HW4.Rmd

Knit on Save | Knit | Run

Source | Visual | Outline




```
72  We notice that most vehicles' highway mpg are between 15-30 mpg. It's kind of
    like a normal distribution but not close.
73
74  Ex2:
75  ```{r}
76  plot(mpg$hwy, mpg$cty, xlab="highway mpg", ylab="city mpg", pch=19)
77  ```
```



68:1 | Chunk 2 | R Markdown

Console

Environment | History | Connections

Diff | Commit

Staged | Status | Path

131 HW1.Rmd

Files | Plots | Packages | Help | Viewer

Zoom | Export

PSTAT131 - main - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

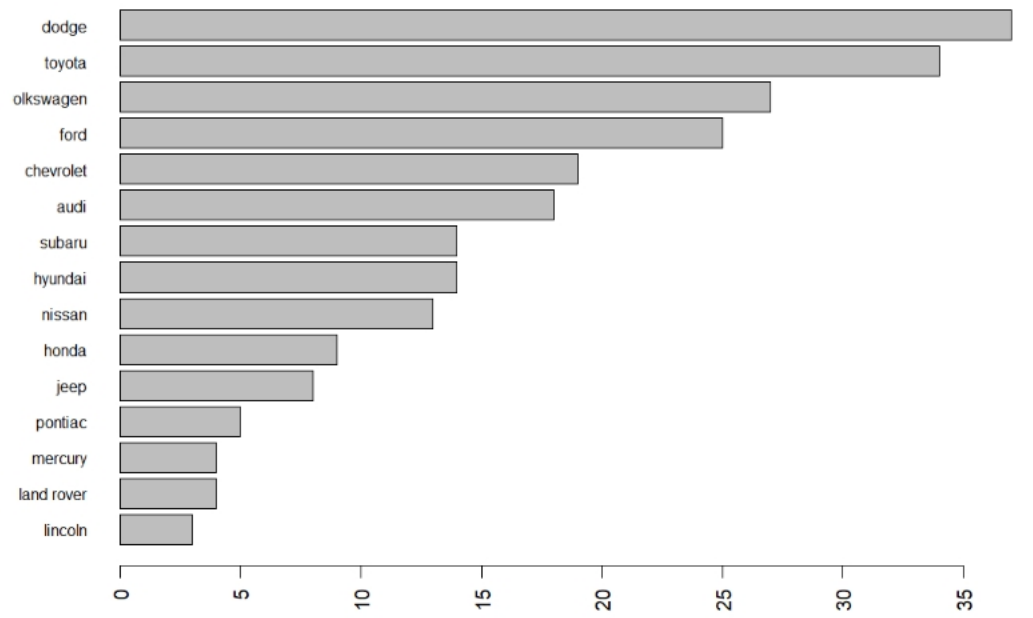
Go to file/function | Addins | PSTAT131

131 HW1.Rmd | 1.R | 126 HW4.Rmd

Knit on Save | Knit | Run

Source | Visual | Outline

```
78 We notice that there is a possible relationship between highway mpg and city
   mpg.
79 That means a car with higher hwy mpg are likely to have higher city mpg. And
   vice versa.
80
81 Ex3:
82 ```{r}
83 count=sort(table(mpg$manufacturer))
84 barplot(count,horiz=TRUE,las=2,cex.names = 0.8)
85 ```
```




```
86 As we can see, Dodge produced the most, and Lincoln produced the least.
87
88
89 Ex4:
90 ```{r}
91 boxplot(hwy~cyl, data=mpg, xlab="Number of Cylinders", ylab="Miles Per
   Gallon(HWY)")
92 ```
```

68:1 | Chunk 2 | R Markdown

Console

Environment | History | Connections

Diff | Commit

Staged | Status | Path

131 HW1.Rmd

Files | Plots | Packages | Help | Viev

Zoom | Export

PSTAT131 - main - RStudio

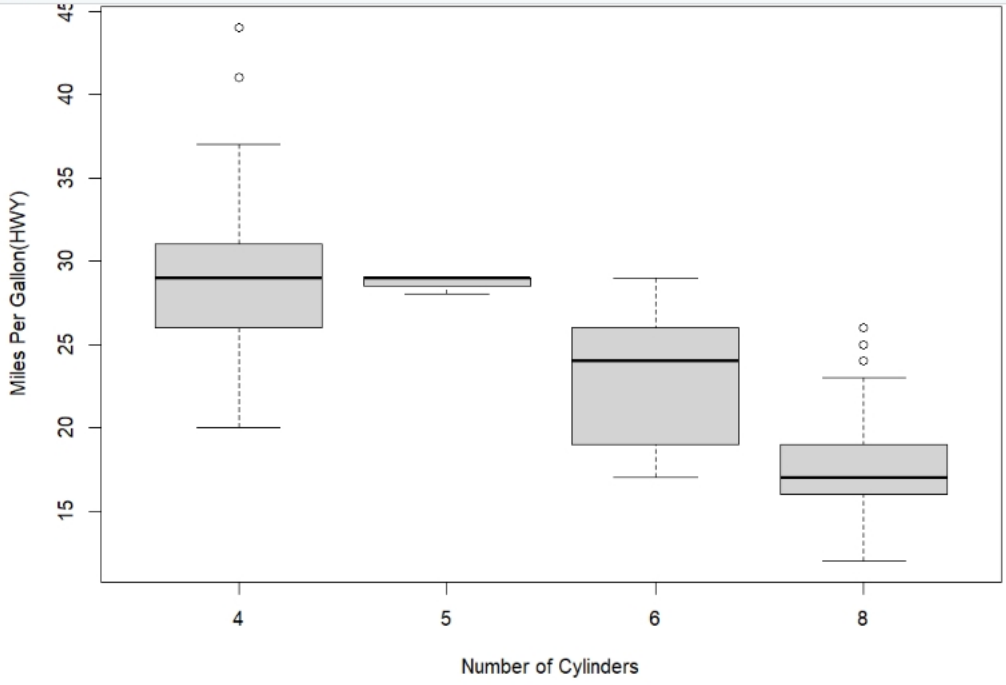
File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | PSTAT131

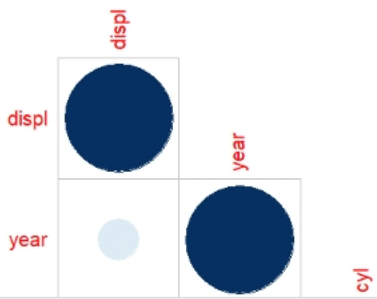
131 HW1.Rmd | 1.R | 126 HW4.Rmd

Knit on Save | Knit | Run

Source | Visual | Outline




```
93  Cars with less cylinders tend to have higher mpg. We can say that cylinder
    number and hwy mpg are negatively correlated.
94
95
96  Ex5:
97  ```{r}
98  install.packages('corrplot')
99  library(corrplot)
100 newmpg=subset(mpg,select=-c(manufacturer, drv, model, trans, fl, class))
101 cormpg=cor(newmpg)
102 corrplot(cormpg,type="lower")
103 ```
```



68:1 | Chunk 2 | R Markdown

Console

Environment | History | Connections

Diff | Commit

Staged | Status | Path

131 HW1.Rmd

Files | Plots | Packages | Help | Viev

Zoom | Export

PSTAT131 - main - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | PSTAT131

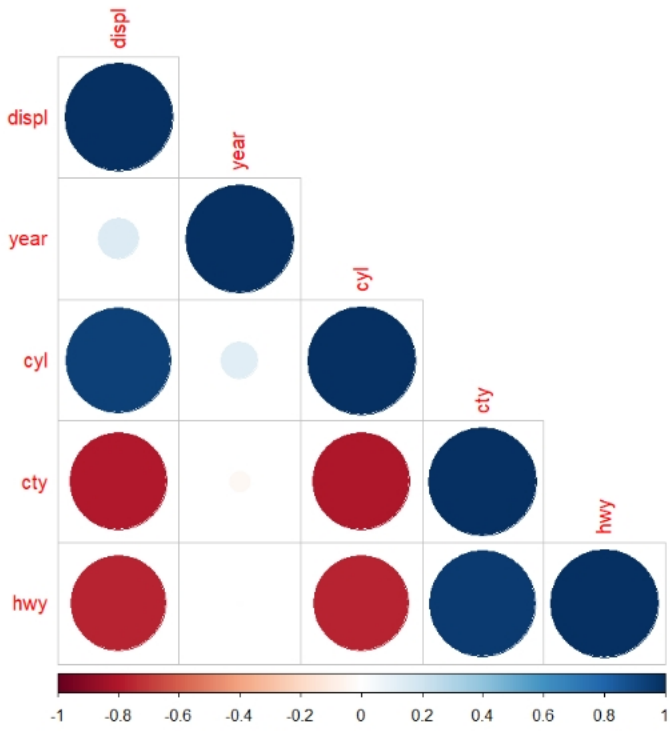
131 HW1.Rmd | 1.R | 126 HW4.Rmd

Knit on Save | Knit | Run

Source | Visual | Outline

Number of Cylinders

```
93  Cars with less cylinders tend to have higher mpg. We can say that cylinder
    number and hwy mpg are negatively correlated.
94
95
96  Ex5:
97  ```{r}
98  install.packages('corrplot')
99  library(corrplot)
100 newmpg=subset(mpg,select=-c(manufacturer, drv, model, trans, fl, class))
101 cormpg=cor(newmpg)
102 corrplot(cormpg,type="lower")
103 ```
```




```
104 For example, displacement is positively correlated to cylinder numbers, city mpg
    is positively correlated to highway mpg, and cylinder number is negatively
    correlated to city or highway mpg. Those relationship make sense to me. Non of
    them is surprising.
105
```

68:1 | Chunk 2 | R Markdown

Console

Environment | History | Connections

Diff | Commit

Staged | Status | Path

131 HW1.Rmd

Files | Plots | Packages | Help | Viev

Zoom | Export