Week 3 Summary

Advait Ashtikar

Table of contents

Tuesday, Jan 24

> **!** TIL
>
> Today, I learnt the following concepts in class:
>
> 1. 'read.csv' and its uses for reading datasets
> 2. 'dplyr' and its use in manipulating datasets
> 3. 'ggplot2' used for plotting graphs

Load Libraries

```
library(dplyr)
library(purrr)
library(tidyverse)
library(ggplot2)
library(ggThemeAssist)
library(forcats)
```

read.csv

1. 'read.csv' is a special case of a function 'read.table'
2. Other more advanced and efficient methods are:
    1. 'read_csv' from 'tidyverse'
    2. 'data.table' package in R

Reading a .csv file from memory:

```
file_location <- "./data.csv"
data_from_csv <- read.csv(file_location)
```

```
Warning in read.table(file = file, header = header, sep = sep, quote = quote, :
incomplete final line found by readTableHeader on './data.csv'
```

```
data_from_csv %>% knitr::kable()
```

Name	Age	Height
Alice	21	5.5
Bob	25	6.2
Charlie	35	5.9

dplyr

After we get the data set we start analyzing the data using 'dplyr' and 'ggplot2'. This is known as **Exploratory Data Analysis**

'dplyr' is used to provide a set of "Verbs" for manipulating data

Q. What makes a dataset clean?

Ans. 1. Get rid of 'Null', 'NA', 'NaN' and 'missing' entries

2. Making sure that all the values for a particular variable are of the same 'data type', **Eg:** 'double', 'character', or 'logical'
3. Each case should be a different value

In general we want the following:

1. Every variable should have its own value
2. Every observation should have its own value
3. Every cell, should have a unique value

Examples of Verbs:

- 'dplyr::summary()', 'dplyr::mutate()'
- 'tidyr::pivot_longer()', 'tidyr::pivot_wider()'
- 'left.join', 'right.join', 'inner.join', 'outer.join'
- left_join', 'right_join', 'inner_join', 'outer_join'

'dplyr' working example using Cars (mpg) dataset

```
head(mpg,10) %>% knitr::kable()
```

manufacturer	model	displ	year	cyl	trans	drv	cty	hwy	fl	class
audi	a4	1.8	1999	4	auto(l5)	f	18	29	p	compact
audi	a4	1.8	1999	4	manual(m5)	f	21	29	p	compact
audi	a4	2.0	2008	4	manual(m6)	f	20	31	p	compact
audi	a4	2.0	2008	4	auto(av)	f	21	30	p	compact
audi	a4	2.8	1999	6	auto(l5)	f	16	26	p	compact
audi	a4	2.8	1999	6	manual(m5)	f	18	26	p	compact
audi	a4	3.1	2008	6	auto(av)	f	18	27	p	compact
audi	a4 quattro	1.8	1999	4	manual(m5)	4	18	26	p	compact
audi	a4 quattro	1.8	1999	4	auto(l5)	4	16	25	p	compact
audi	a4 quattro	2.0	2008	4	manual(m6)	4	20	28	p	compact

'dplyr' working example using Iris (flower petal) dataset

```
head(iris, 5) %>% knitr::kable()
```

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa

Examples of different functions:

1. Select - selects a subset of the columns

```
mpg %>%
  select(c(model, displ, class)) %>%
  knitr::kable()
```

model	displ	class
a4	1.8	compact
a4	1.8	compact
a4	2.0	compact
a4	2.0	compact
a4	2.8	compact
a4	2.8	compact
a4	3.1	compact
a4 quattro	1.8	compact
a4 quattro	1.8	compact
a4 quattro	2.0	compact
a4 quattro	2.0	compact
a4 quattro	2.8	compact
a4 quattro	2.8	compact
a4 quattro	3.1	compact
a4 quattro	3.1	compact
a6 quattro	2.8	midsize
a6 quattro	3.1	midsize
a6 quattro	4.2	midsize
c1500 suburban 2wd	5.3	suv
c1500 suburban 2wd	5.3	suv

model	displ	class
c1500 suburban 2wd	5.3	suv
c1500 suburban 2wd	5.7	suv
c1500 suburban 2wd	6.0	suv
corvette	5.7	2seater
corvette	5.7	2seater
corvette	6.2	2seater
corvette	6.2	2seater
corvette	7.0	2seater
k1500 tahoe 4wd	5.3	suv
k1500 tahoe 4wd	5.3	suv
k1500 tahoe 4wd	5.7	suv
k1500 tahoe 4wd	6.5	suv
malibu	2.4	midsize
malibu	2.4	midsize
malibu	3.1	midsize
malibu	3.5	midsize
malibu	3.6	midsize
caravan 2wd	2.4	minivan
caravan 2wd	3.0	minivan
caravan 2wd	3.3	minivan
caravan 2wd	3.3	minivan
caravan 2wd	3.3	minivan
caravan 2wd	3.3	minivan
caravan 2wd	3.3	minivan
caravan 2wd	3.8	minivan
caravan 2wd	3.8	minivan
caravan 2wd	3.8	minivan
caravan 2wd	4.0	minivan
dakota pickup 4wd	3.7	pickup
dakota pickup 4wd	3.7	pickup
dakota pickup 4wd	3.9	pickup
dakota pickup 4wd	3.9	pickup
dakota pickup 4wd	4.7	pickup
dakota pickup 4wd	4.7	pickup
dakota pickup 4wd	4.7	pickup
dakota pickup 4wd	5.2	pickup
dakota pickup 4wd	5.2	pickup
durango 4wd	3.9	suv
durango 4wd	4.7	suv
durango 4wd	4.7	suv
durango 4wd	4.7	suv

model	displ	class
durango 4wd	5.2	suv
durango 4wd	5.7	suv
durango 4wd	5.9	suv
ram 1500 pickup 4wd	4.7	pickup
ram 1500 pickup 4wd	4.7	pickup
ram 1500 pickup 4wd	4.7	pickup
ram 1500 pickup 4wd	4.7	pickup
ram 1500 pickup 4wd	4.7	pickup
ram 1500 pickup 4wd	4.7	pickup
ram 1500 pickup 4wd	5.2	pickup
ram 1500 pickup 4wd	5.2	pickup
ram 1500 pickup 4wd	5.7	pickup
ram 1500 pickup 4wd	5.9	pickup
expedition 2wd	4.6	suv
expedition 2wd	5.4	suv
expedition 2wd	5.4	suv
explorer 4wd	4.0	suv
explorer 4wd	4.0	suv
explorer 4wd	4.0	suv
explorer 4wd	4.0	suv
explorer 4wd	4.6	suv
explorer 4wd	5.0	suv
f150 pickup 4wd	4.2	pickup
f150 pickup 4wd	4.2	pickup
f150 pickup 4wd	4.6	pickup
f150 pickup 4wd	4.6	pickup
f150 pickup 4wd	4.6	pickup
f150 pickup 4wd	5.4	pickup
f150 pickup 4wd	5.4	pickup
mustang	3.8	subcompact
mustang	3.8	subcompact
mustang	4.0	subcompact
mustang	4.0	subcompact
mustang	4.6	subcompact
mustang	4.6	subcompact
mustang	4.6	subcompact
mustang	4.6	subcompact
mustang	5.4	subcompact
civic	1.6	subcompact
civic	1.6	subcompact
civic	1.6	subcompact

model	displ	class
civic	1.6	subcompact
civic	1.6	subcompact
civic	1.8	subcompact
civic	1.8	subcompact
civic	1.8	subcompact
civic	2.0	subcompact
sonata	2.4	midsize
sonata	2.4	midsize
sonata	2.4	midsize
sonata	2.4	midsize
sonata	2.5	midsize
sonata	2.5	midsize
sonata	3.3	midsize
tiburon	2.0	subcompact
tiburon	2.0	subcompact
tiburon	2.0	subcompact
tiburon	2.0	subcompact
tiburon	2.7	subcompact
tiburon	2.7	subcompact
tiburon	2.7	subcompact
grand cherokee 4wd	3.0	suv
grand cherokee 4wd	3.7	suv
grand cherokee 4wd	4.0	suv
grand cherokee 4wd	4.7	suv
grand cherokee 4wd	4.7	suv
grand cherokee 4wd	4.7	suv
grand cherokee 4wd	5.7	suv
grand cherokee 4wd	6.1	suv
range rover	4.0	suv
range rover	4.2	suv
range rover	4.4	suv
range rover	4.6	suv
navigator 2wd	5.4	suv
navigator 2wd	5.4	suv
navigator 2wd	5.4	suv
mountaineer 4wd	4.0	suv
mountaineer 4wd	4.0	suv
mountaineer 4wd	4.6	suv
mountaineer 4wd	5.0	suv
altima	2.4	compact
altima	2.4	compact

model	displ	class
altima	2.5	midsize
altima	2.5	midsize
altima	3.5	midsize
altima	3.5	midsize
maxima	3.0	midsize
maxima	3.0	midsize
maxima	3.5	midsize
pathfinder 4wd	3.3	suv
pathfinder 4wd	3.3	suv
pathfinder 4wd	4.0	suv
pathfinder 4wd	5.6	suv
grand prix	3.1	midsize
grand prix	3.8	midsize
grand prix	3.8	midsize
grand prix	3.8	midsize
grand prix	5.3	midsize
forester awd	2.5	suv
forester awd	2.5	suv
forester awd	2.5	suv
forester awd	2.5	suv
forester awd	2.5	suv
forester awd	2.5	suv
impreza awd	2.2	subcompact
impreza awd	2.2	subcompact
impreza awd	2.5	subcompact
impreza awd	2.5	subcompact
impreza awd	2.5	compact
impreza awd	2.5	compact
impreza awd	2.5	compact
impreza awd	2.5	compact
4runner 4wd	2.7	suv
4runner 4wd	2.7	suv
4runner 4wd	3.4	suv
4runner 4wd	3.4	suv
4runner 4wd	4.0	suv
4runner 4wd	4.7	suv
camry	2.2	midsize
camry	2.2	midsize
camry	2.4	midsize
camry	2.4	midsize
camry	3.0	midsize

model	displ	class
camry	3.0	midsize
camry	3.5	midsize
camry solara	2.2	compact
camry solara	2.2	compact
camry solara	2.4	compact
camry solara	2.4	compact
camry solara	3.0	compact
camry solara	3.0	compact
camry solara	3.3	compact
corolla	1.8	compact
corolla	1.8	compact
corolla	1.8	compact
corolla	1.8	compact
corolla	1.8	compact
land cruiser wagon 4wd	4.7	suv
land cruiser wagon 4wd	5.7	suv
toyota tacoma 4wd	2.7	pickup
toyota tacoma 4wd	2.7	pickup
toyota tacoma 4wd	2.7	pickup
toyota tacoma 4wd	3.4	pickup
toyota tacoma 4wd	3.4	pickup
toyota tacoma 4wd	4.0	pickup
toyota tacoma 4wd	4.0	pickup
gti	2.0	compact
gti	2.0	compact
gti	2.0	compact
gti	2.0	compact
gti	2.8	compact
jetta	1.9	compact
jetta	2.0	compact
jetta	2.0	compact
jetta	2.0	compact
jetta	2.0	compact
jetta	2.5	compact
jetta	2.5	compact
jetta	2.8	compact
jetta	2.8	compact
new beetle	1.9	subcompact
new beetle	1.9	subcompact
new beetle	2.0	subcompact
new beetle	2.0	subcompact

model	displ	class
new beetle	2.5	subcompact
new beetle	2.5	subcompact
passat	1.8	midsize
passat	1.8	midsize
passat	2.0	midsize
passat	2.0	midsize
passat	2.8	midsize
passat	2.8	midsize
passat	3.6	midsize

2. Mutate - creates new columns from existing rows

```
iris %>%
  mutate(Sepal_Area = Sepal.Length * Sepal.Width) %>%
  knitr::kable()
```

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species	Sepal_Area
5.1	3.5	1.4	0.2	setosa	17.85
4.9	3.0	1.4	0.2	setosa	14.70
4.7	3.2	1.3	0.2	setosa	15.04
4.6	3.1	1.5	0.2	setosa	14.26
5.0	3.6	1.4	0.2	setosa	18.00
5.4	3.9	1.7	0.4	setosa	21.06
4.6	3.4	1.4	0.3	setosa	15.64
5.0	3.4	1.5	0.2	setosa	17.00
4.4	2.9	1.4	0.2	setosa	12.76
4.9	3.1	1.5	0.1	setosa	15.19
5.4	3.7	1.5	0.2	setosa	19.98
4.8	3.4	1.6	0.2	setosa	16.32
4.8	3.0	1.4	0.1	setosa	14.40
4.3	3.0	1.1	0.1	setosa	12.90
5.8	4.0	1.2	0.2	setosa	23.20
5.7	4.4	1.5	0.4	setosa	25.08
5.4	3.9	1.3	0.4	setosa	21.06
5.1	3.5	1.4	0.3	setosa	17.85
5.7	3.8	1.7	0.3	setosa	21.66
5.1	3.8	1.5	0.3	setosa	19.38
5.4	3.4	1.7	0.2	setosa	18.36
5.1	3.7	1.5	0.4	setosa	18.87

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species	Sepal_Area
4.6	3.6	1.0	0.2	setosa	16.56
5.1	3.3	1.7	0.5	setosa	16.83
4.8	3.4	1.9	0.2	setosa	16.32
5.0	3.0	1.6	0.2	setosa	15.00
5.0	3.4	1.6	0.4	setosa	17.00
5.2	3.5	1.5	0.2	setosa	18.20
5.2	3.4	1.4	0.2	setosa	17.68
4.7	3.2	1.6	0.2	setosa	15.04
4.8	3.1	1.6	0.2	setosa	14.88
5.4	3.4	1.5	0.4	setosa	18.36
5.2	4.1	1.5	0.1	setosa	21.32
5.5	4.2	1.4	0.2	setosa	23.10
4.9	3.1	1.5	0.2	setosa	15.19
5.0	3.2	1.2	0.2	setosa	16.00
5.5	3.5	1.3	0.2	setosa	19.25
4.9	3.6	1.4	0.1	setosa	17.64
4.4	3.0	1.3	0.2	setosa	13.20
5.1	3.4	1.5	0.2	setosa	17.34
5.0	3.5	1.3	0.3	setosa	17.50
4.5	2.3	1.3	0.3	setosa	10.35
4.4	3.2	1.3	0.2	setosa	14.08
5.0	3.5	1.6	0.6	setosa	17.50
5.1	3.8	1.9	0.4	setosa	19.38
4.8	3.0	1.4	0.3	setosa	14.40
5.1	3.8	1.6	0.2	setosa	19.38
4.6	3.2	1.4	0.2	setosa	14.72
5.3	3.7	1.5	0.2	setosa	19.61
5.0	3.3	1.4	0.2	setosa	16.50
7.0	3.2	4.7	1.4	versicolor	22.40
6.4	3.2	4.5	1.5	versicolor	20.48
6.9	3.1	4.9	1.5	versicolor	21.39
5.5	2.3	4.0	1.3	versicolor	12.65
6.5	2.8	4.6	1.5	versicolor	18.20
5.7	2.8	4.5	1.3	versicolor	15.96
6.3	3.3	4.7	1.6	versicolor	20.79
4.9	2.4	3.3	1.0	versicolor	11.76
6.6	2.9	4.6	1.3	versicolor	19.14
5.2	2.7	3.9	1.4	versicolor	14.04
5.0	2.0	3.5	1.0	versicolor	10.00
5.9	3.0	4.2	1.5	versicolor	17.70
6.0	2.2	4.0	1.0	versicolor	13.20

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species	Sepal_Area
6.1	2.9	4.7	1.4	versicolor	17.69
5.6	2.9	3.6	1.3	versicolor	16.24
6.7	3.1	4.4	1.4	versicolor	20.77
5.6	3.0	4.5	1.5	versicolor	16.80
5.8	2.7	4.1	1.0	versicolor	15.66
6.2	2.2	4.5	1.5	versicolor	13.64
5.6	2.5	3.9	1.1	versicolor	14.00
5.9	3.2	4.8	1.8	versicolor	18.88
6.1	2.8	4.0	1.3	versicolor	17.08
6.3	2.5	4.9	1.5	versicolor	15.75
6.1	2.8	4.7	1.2	versicolor	17.08
6.4	2.9	4.3	1.3	versicolor	18.56
6.6	3.0	4.4	1.4	versicolor	19.80
6.8	2.8	4.8	1.4	versicolor	19.04
6.7	3.0	5.0	1.7	versicolor	20.10
6.0	2.9	4.5	1.5	versicolor	17.40
5.7	2.6	3.5	1.0	versicolor	14.82
5.5	2.4	3.8	1.1	versicolor	13.20
5.5	2.4	3.7	1.0	versicolor	13.20
5.8	2.7	3.9	1.2	versicolor	15.66
6.0	2.7	5.1	1.6	versicolor	16.20
5.4	3.0	4.5	1.5	versicolor	16.20
6.0	3.4	4.5	1.6	versicolor	20.40
6.7	3.1	4.7	1.5	versicolor	20.77
6.3	2.3	4.4	1.3	versicolor	14.49
5.6	3.0	4.1	1.3	versicolor	16.80
5.5	2.5	4.0	1.3	versicolor	13.75
5.5	2.6	4.4	1.2	versicolor	14.30
6.1	3.0	4.6	1.4	versicolor	18.30
5.8	2.6	4.0	1.2	versicolor	15.08
5.0	2.3	3.3	1.0	versicolor	11.50
5.6	2.7	4.2	1.3	versicolor	15.12
5.7	3.0	4.2	1.2	versicolor	17.10
5.7	2.9	4.2	1.3	versicolor	16.53
6.2	2.9	4.3	1.3	versicolor	17.98
5.1	2.5	3.0	1.1	versicolor	12.75
5.7	2.8	4.1	1.3	versicolor	15.96
6.3	3.3	6.0	2.5	virginica	20.79
5.8	2.7	5.1	1.9	virginica	15.66
7.1	3.0	5.9	2.1	virginica	21.30
6.3	2.9	5.6	1.8	virginica	18.27

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species	Sepal_Area
6.5	3.0	5.8	2.2	virginica	19.50
7.6	3.0	6.6	2.1	virginica	22.80
4.9	2.5	4.5	1.7	virginica	12.25
7.3	2.9	6.3	1.8	virginica	21.17
6.7	2.5	5.8	1.8	virginica	16.75
7.2	3.6	6.1	2.5	virginica	25.92
6.5	3.2	5.1	2.0	virginica	20.80
6.4	2.7	5.3	1.9	virginica	17.28
6.8	3.0	5.5	2.1	virginica	20.40
5.7	2.5	5.0	2.0	virginica	14.25
5.8	2.8	5.1	2.4	virginica	16.24
6.4	3.2	5.3	2.3	virginica	20.48
6.5	3.0	5.5	1.8	virginica	19.50
7.7	3.8	6.7	2.2	virginica	29.26
7.7	2.6	6.9	2.3	virginica	20.02
6.0	2.2	5.0	1.5	virginica	13.20
6.9	3.2	5.7	2.3	virginica	22.08
5.6	2.8	4.9	2.0	virginica	15.68
7.7	2.8	6.7	2.0	virginica	21.56
6.3	2.7	4.9	1.8	virginica	17.01
6.7	3.3	5.7	2.1	virginica	22.11
7.2	3.2	6.0	1.8	virginica	23.04
6.2	2.8	4.8	1.8	virginica	17.36
6.1	3.0	4.9	1.8	virginica	18.30
6.4	2.8	5.6	2.1	virginica	17.92
7.2	3.0	5.8	1.6	virginica	21.60
7.4	2.8	6.1	1.9	virginica	20.72
7.9	3.8	6.4	2.0	virginica	30.02
6.4	2.8	5.6	2.2	virginica	17.92
6.3	2.8	5.1	1.5	virginica	17.64
6.1	2.6	5.6	1.4	virginica	15.86
7.7	3.0	6.1	2.3	virginica	23.10
6.3	3.4	5.6	2.4	virginica	21.42
6.4	3.1	5.5	1.8	virginica	19.84
6.0	3.0	4.8	1.8	virginica	18.00
6.9	3.1	5.4	2.1	virginica	21.39
6.7	3.1	5.6	2.4	virginica	20.77
6.9	3.1	5.1	2.3	virginica	21.39
5.8	2.7	5.1	1.9	virginica	15.66
6.8	3.2	5.9	2.3	virginica	21.76
6.7	3.3	5.7	2.5	virginica	22.11

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species	Sepal_Area
6.7	3.0	5.2	2.3	virginica	20.10
6.3	2.5	5.0	1.9	virginica	15.75
6.5	3.0	5.2	2.0	virginica	19.50
6.2	3.4	5.4	2.3	virginica	21.08
5.9	3.0	5.1	1.8	virginica	17.70

3. Filter -

```
mpg %>%
  filter(class == "compact") %>%
  knitr::kable()
```

manufacturer	model	displ	year	cyl	trans	drv	cty	hwy	fl	class
audi	a4	1.8	1999	4	auto(l5)	f	18	29	p	compact
audi	a4	1.8	1999	4	manual(m5)	f	21	29	p	compact
audi	a4	2.0	2008	4	manual(m6)	f	20	31	p	compact
audi	a4	2.0	2008	4	auto(av)	f	21	30	p	compact
audi	a4	2.8	1999	6	auto(l5)	f	16	26	p	compact
audi	a4	2.8	1999	6	manual(m5)	f	18	26	p	compact
audi	a4	3.1	2008	6	auto(av)	f	18	27	p	compact
audi	a4 quattro	1.8	1999	4	manual(m5)	4	18	26	p	compact
audi	a4 quattro	1.8	1999	4	auto(l5)	4	16	25	p	compact
audi	a4 quattro	2.0	2008	4	manual(m6)	4	20	28	p	compact
audi	a4 quattro	2.0	2008	4	auto(s6)	4	19	27	p	compact
audi	a4 quattro	2.8	1999	6	auto(l5)	4	15	25	p	compact
audi	a4 quattro	2.8	1999	6	manual(m5)	4	17	25	p	compact
audi	a4 quattro	3.1	2008	6	auto(s6)	4	17	25	p	compact
audi	a4 quattro	3.1	2008	6	manual(m6)	4	15	25	p	compact
nissan	altima	2.4	1999	4	manual(m5)	f	21	29	r	compact
nissan	altima	2.4	1999	4	auto(l4)	f	19	27	r	compact
subaru	impreza awd	2.5	2008	4	auto(s4)	4	20	25	p	compact
subaru	impreza awd	2.5	2008	4	auto(s4)	4	20	27	r	compact
subaru	impreza awd	2.5	2008	4	manual(m5)	4	19	25	p	compact
subaru	impreza awd	2.5	2008	4	manual(m5)	4	20	27	r	compact
toyota	camry solara	2.2	1999	4	auto(l4)	f	21	27	r	compact

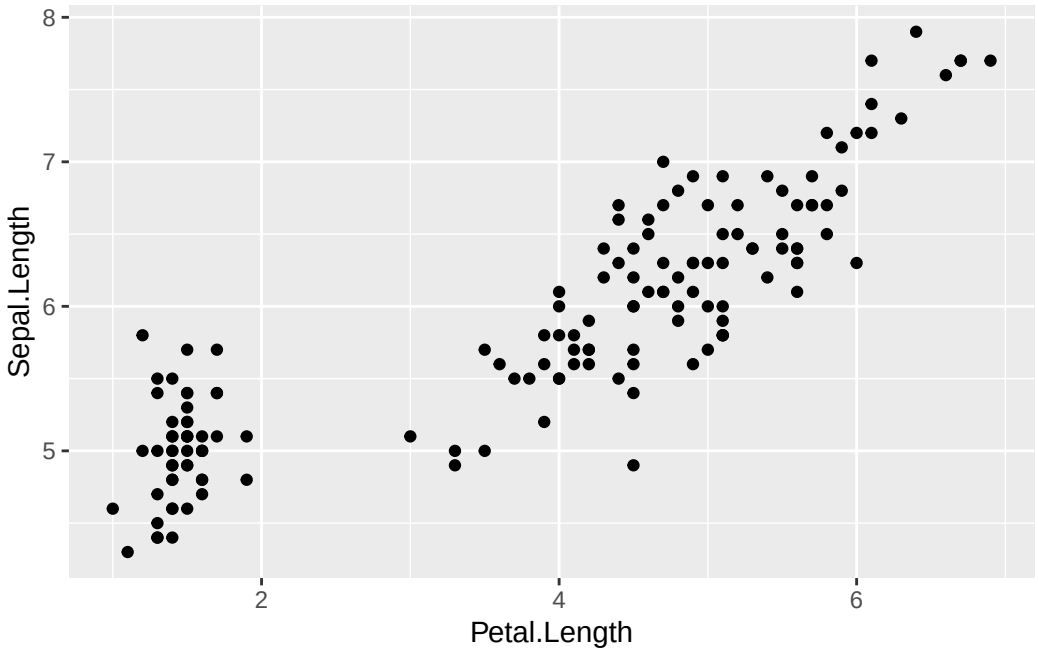
manufacturer	model	displ	year	cyl	trans	drv	cty	hwy	fl	class
toyota	camry solara	2.2	1999	4	manual(m5)	f	21	29	r	compact
toyota	camry solara	2.4	2008	4	manual(m5)	f	21	31	r	compact
toyota	camry solara	2.4	2008	4	auto(s5)	f	22	31	r	compact
toyota	camry solara	3.0	1999	6	auto(l4)	f	18	26	r	compact
toyota	camry solara	3.0	1999	6	manual(m5)	f	18	26	r	compact
toyota	camry solara	3.3	2008	6	auto(s5)	f	18	27	r	compact
toyota	corolla	1.8	1999	4	auto(l3)	f	24	30	r	compact
toyota	corolla	1.8	1999	4	auto(l4)	f	24	33	r	compact
toyota	corolla	1.8	1999	4	manual(m5)	f	26	35	r	compact
toyota	corolla	1.8	2008	4	manual(m5)	f	28	37	r	compact
toyota	corolla	1.8	2008	4	auto(l4)	f	26	35	r	compact
volkswagen	gti	2.0	1999	4	manual(m5)	f	21	29	r	compact
volkswagen	gti	2.0	1999	4	auto(l4)	f	19	26	r	compact
volkswagen	gti	2.0	2008	4	manual(m6)	f	21	29	p	compact
volkswagen	gti	2.0	2008	4	auto(s6)	f	22	29	p	compact
volkswagen	gti	2.8	1999	6	manual(m5)	f	17	24	r	compact
volkswagen	jetta	1.9	1999	4	manual(m5)	f	33	44	d	compact
volkswagen	jetta	2.0	1999	4	manual(m5)	f	21	29	r	compact
volkswagen	jetta	2.0	1999	4	auto(l4)	f	19	26	r	compact
volkswagen	jetta	2.0	2008	4	auto(s6)	f	22	29	p	compact
volkswagen	jetta	2.0	2008	4	manual(m6)	f	21	29	p	compact
volkswagen	jetta	2.5	2008	5	auto(s6)	f	21	29	r	compact
volkswagen	jetta	2.5	2008	5	manual(m5)	f	21	29	r	compact
volkswagen	jetta	2.8	1999	6	auto(l4)	f	16	23	r	compact
volkswagen	jetta	2.8	1999	6	manual(m5)	f	17	24	r	compact

ggplot2

The 'gg' in 'ggplot2' stands for **G**rammar for **G**raphics

Examples of ggplot2

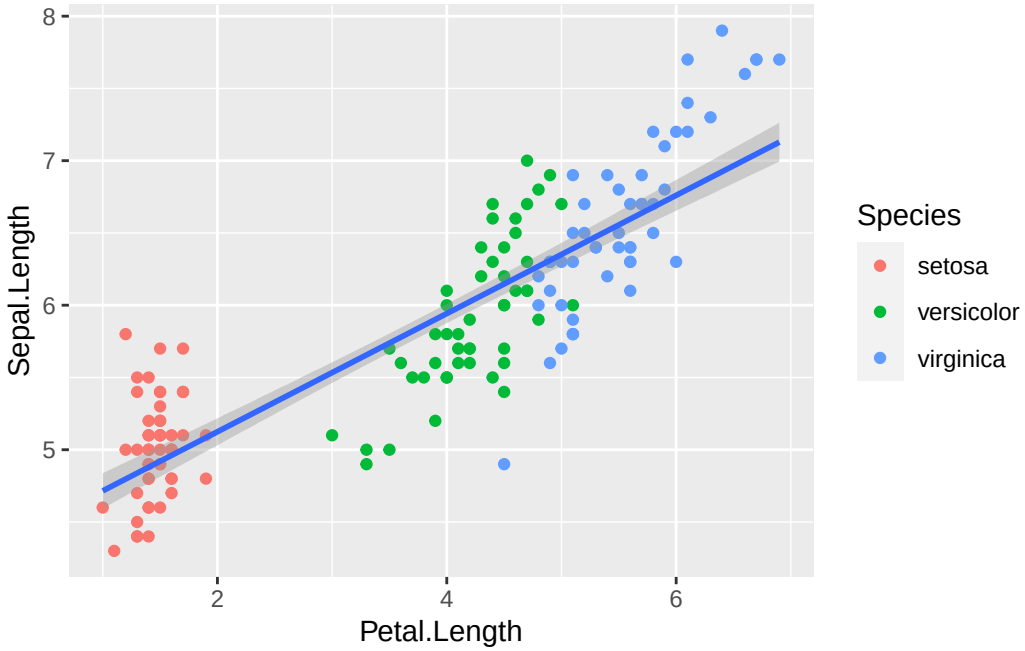
```
library(ggplot2)
plt <- ggplot(iris)
#To add points to graph
plt + geom_point(aes(x=Petal.Length, y=Sepal.Length))
```



If we wish to add trend lines to the points on the graph:

```
plt + geom_point(aes(x=Petal.Length, y=Sepal.Length, color=Species)) + geom_smooth(aes(x=P
```

```
`geom_smooth()` using formula = 'y ~ x'
```



Thursday, Jan 26

> **!** TIL
>
> Include a *very brief* summary of what you learnt in this class here.
> Today, I learnt the following concepts in class:
>
> 1. ggThemeAssist
> 2. More on data types
> 3. Factors
> 4. purrr

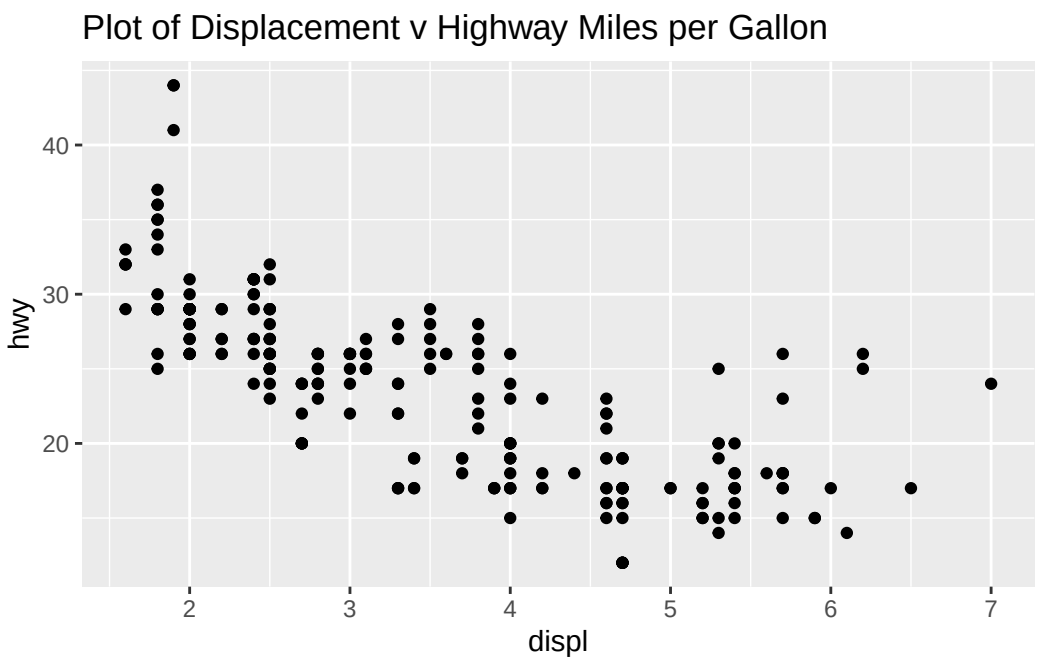
ggThemeAssist

'ggThemeAssist' helps in easily modifying different characteristics of a plot.

```
library(ggThemeAssist)
library(ggplot2)

ggplot(mpg) +
```

```
geom_point(aes(x=displ, y=hwy))+labs(title = "Plot of Displacement v Highway Miles per G
```



Data Types

1. String, Eg: 'r x <- "this is a character"; x'
2. Integer, Eg: '[1, 2, 3]'
3. Double, Eg: '[2.3, 3.14159, 0.9]'
4. Booleans, Eg: 'TRUE/FALSE'

What are factors?

- Factors are categorical variables
- 'var' contains the country code for people in North America

```
var <- c(
  "USA",
  "USA",
  "CAN",
  "CAN",
  "CAN",
```

```
  "CAN",
  "MEX",
  "MEX"
)
```

To tell 'R' that this is explicitly categorcial and not just a vector of strings, you have to specify the following:

```
as.factor(var)
```

```
[1] USA USA CAN CAN CAN CAN MEX MEX
Levels: CAN MEX USA
```

Another Example:

```
head(iris, 3) %>%
  knitr::kable()
```

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa

```
iris$Species
```

```
  [1] setosa     setosa     setosa     setosa     setosa     setosa
  [7] setosa     setosa     setosa     setosa     setosa     setosa
 [13] setosa     setosa     setosa     setosa     setosa     setosa
 [19] setosa     setosa     setosa     setosa     setosa     setosa
 [25] setosa     setosa     setosa     setosa     setosa     setosa
 [31] setosa     setosa     setosa     setosa     setosa     setosa
 [37] setosa     setosa     setosa     setosa     setosa     setosa
 [43] setosa     setosa     setosa     setosa     setosa     setosa
 [49] setosa     setosa     versicolor versicolor versicolor versicolor
 [55] versicolor versicolor versicolor versicolor versicolor versicolor
 [61] versicolor versicolor versicolor versicolor versicolor versicolor
 [67] versicolor versicolor versicolor versicolor versicolor versicolor
 [73] versicolor versicolor versicolor versicolor versicolor versicolor
 [79] versicolor versicolor versicolor versicolor versicolor versicolor
```

```
 [85] versicolor versicolor versicolor versicolor versicolor versicolor
 [91] versicolor versicolor versicolor versicolor versicolor versicolor
 [97] versicolor versicolor versicolor versicolor virginica  virginica
[103] virginica  virginica  virginica  virginica  virginica  virginica
[109] virginica  virginica  virginica  virginica  virginica  virginica
[115] virginica  virginica  virginica  virginica  virginica  virginica
[121] virginica  virginica  virginica  virginica  virginica  virginica
[127] virginica  virginica  virginica  virginica  virginica  virginica
[133] virginica  virginica  virginica  virginica  virginica  virginica
[139] virginica  virginica  virginica  virginica  virginica  virginica
[145] virginica  virginica  virginica  virginica  virginica  virginica
Levels: setosa versicolor virginica
```

Similarly, if we look at 'mpg'

```
head(mpg, 3) %>% knitr::kable()
```

manufacturer	model	displ	year	cyl	trans	drv	cty	hwy	fl	class
audi	a4	1.8	1999	4	auto(l5)	f	18	29	p	compact
audi	a4	1.8	1999	4	manual(m5)	f	21	29	p	compact
audi	a4	2.0	2008	4	manual(m6)	f	20	31	p	compact

Lets have a look at 'class'

```
as.factor(mpg$class)
```

```
 [1] compact    compact    compact    compact    compact    compact
 [7] compact    compact    compact    compact    compact    compact
[13] compact    compact    compact    midsize    midsize    midsize
[19] suv        suv        suv        suv        suv        2seater
[25] 2seater    2seater    2seater    2seater    suv        suv
[31] suv        suv        midsize    midsize    midsize    midsize
[37] midsize    minivan    minivan    minivan    minivan    minivan
[43] minivan    minivan    minivan    minivan    minivan    minivan
[49] pickup     pickup     pickup     pickup     pickup     pickup
[55] pickup     pickup     pickup     suv        suv        suv
[61] suv        suv        suv        suv        pickup     pickup
[67] pickup     pickup     pickup     pickup     pickup     pickup
[73] pickup     pickup     suv        suv        suv        suv
[79] suv        suv        suv        suv        suv        pickup
```

```
 [85] pickup     pickup     pickup     pickup     pickup     pickup
 [91] subcompact subcompact subcompact subcompact subcompact subcompact
 [97] subcompact subcompact subcompact subcompact subcompact subcompact
[103] subcompact subcompact subcompact subcompact subcompact subcompact
[109] midsize    midsize    midsize    midsize    midsize    midsize
[115] midsize    subcompact subcompact subcompact subcompact subcompact
[121] subcompact subcompact suv         suv         suv         suv
[127] suv        suv        suv         suv         suv         suv
[133] suv        suv        suv         suv         suv         suv
[139] suv        suv        suv         compact     compact     midsize
[145] midsize    midsize    midsize     midsize     midsize     midsize
[151] suv        suv        suv         suv         midsize     midsize
[157] midsize    midsize    midsize     suv         suv         suv
[163] suv        suv        suv         subcompact  subcompact  subcompact
[169] subcompact compact    compact     compact     compact     suv
[175] suv        suv        suv         suv         suv         midsize
[181] midsize    midsize    midsize     midsize     midsize     midsize
[187] compact    compact    compact     compact     compact     compact
[193] compact    compact    compact     compact     compact     compact
[199] suv        suv        pickup      pickup      pickup      pickup
[205] pickup     pickup     pickup      compact     compact     compact
[211] compact    compact    compact     compact     compact     compact
[217] compact    compact    compact     compact     compact     subcompact
[223] subcompact subcompact subcompact  subcompact  subcompact  midsize
[229] midsize    midsize    midsize     midsize     midsize     midsize
Levels: 2seater compact midsize minivan pickup subcompact suv
```

This is where the 'forcats' package is really useful:

```
library(forcats)
manufacturer <- as.factor(mpg$manufacturer)
fct_reorder(manufacturer, mpg$hwy, min)
```

```
 [1] audi       audi       audi       audi       audi       audi
 [7] audi       audi       audi       audi       audi       audi
[13] audi       audi       audi       audi       audi       audi
[19] chevrolet  chevrolet  chevrolet  chevrolet  chevrolet  chevrolet
[25] chevrolet  chevrolet  chevrolet  chevrolet  chevrolet  chevrolet
[31] chevrolet  chevrolet  chevrolet  chevrolet  chevrolet  chevrolet
[37] chevrolet  dodge      dodge      dodge      dodge      dodge
[43] dodge      dodge      dodge      dodge      dodge      dodge
[49] dodge      dodge      dodge      dodge      dodge      dodge
```

```
 [55] dodge      dodge      dodge      dodge      dodge      dodge
 [61] dodge      dodge      dodge      dodge      dodge      dodge
 [67] dodge      dodge      dodge      dodge      dodge      dodge
 [73] dodge      dodge      ford       ford       ford       ford
 [79] ford       ford       ford       ford       ford       ford
 [85] ford       ford       ford       ford       ford       ford
 [91] ford       ford       ford       ford       ford       ford
 [97] ford       ford       ford       honda      honda      honda
[103] honda      honda      honda      honda      honda      honda
[109] hyundai    hyundai    hyundai    hyundai    hyundai    hyundai
[115] hyundai    hyundai    hyundai    hyundai    hyundai    hyundai
[121] hyundai    hyundai    jeep       jeep       jeep       jeep
[127] jeep       jeep       jeep       jeep       land rover land rover
[133] land rover land rover lincoln    lincoln    lincoln    mercury
[139] mercury    mercury    mercury    nissan     nissan     nissan
[145] nissan     nissan     nissan     nissan     nissan     nissan
[151] nissan     nissan     nissan     nissan     pontiac    pontiac
[157] pontiac    pontiac    pontiac    subaru     subaru     subaru
[163] subaru     subaru     subaru     subaru     subaru     subaru
[169] subaru     subaru     subaru     subaru     subaru     toyota
[175] toyota     toyota     toyota     toyota     toyota     toyota
[181] toyota     toyota     toyota     toyota     toyota     toyota
[187] toyota     toyota     toyota     toyota     toyota     toyota
[193] toyota     toyota     toyota     toyota     toyota     toyota
[199] toyota     toyota     toyota     toyota     toyota     toyota
[205] toyota     toyota     toyota     volkswagen volkswagen volkswagen
[211] volkswagen volkswagen volkswagen volkswagen volkswagen volkswagen
[217] volkswagen volkswagen volkswagen volkswagen volkswagen volkswagen
[223] volkswagen volkswagen volkswagen volkswagen volkswagen volkswagen
[229] volkswagen volkswagen volkswagen volkswagen volkswagen volkswagen
15 Levels: dodge jeep chevrolet ford land rover toyota lincoln ... honda
```

purrr

This package provides a set of functional programming tools. Its best illustrated through an example:

Consider the following procedure: We want to

1. Filter 'iris' by species
2. Compute the 'Sepal.Area' as 'Sepal.Length' $\times$ 'Sepal.Width'
3. Find the average of 'Sepal.Area' for every flower in the species

```
iris %>%
  mutate(Sepal.Area = Sepal.Length * Sepal.Width) %>%
  group_by() %>%
  summarize()
```

```
# A tibble: 1 x 0
```

Consider the following task:

1. Take a number 'i' from $1 \ldots 10$
2. Create a matrix with random entries of dimension 'i' $\times$ 'i'
3. Compute the average of the elements of the matrix
4. Print it

```
results <- c()
for (i in 1:10){
  M <- matrix(
    runif(i*i), nrow=1
  )
  results[i] <- mean(M)
}
results
```

```
 [1] 0.4922043 0.2795106 0.5089657 0.5199897 0.6087588 0.5061262 0.4953501
 [8] 0.4835951 0.4695397 0.5106236
```

A functional way to think of this is as follows:

$i \rightarrow M_{i \times i} \rightarrow mean(M)$$

```
library(purrr)
map(
  1:10,
  function(i){
    mean(
      matrix(
        runif(i * i), nrow=1
      )
    )
  }
)
```

```
[[1]]
[1] 0.6077387

[[2]]
[1] 0.3191795

[[3]]
[1] 0.5999648

[[4]]
[1] 0.6473855

[[5]]
[1] 0.4276901

[[6]]
[1] 0.5208496

[[7]]
[1] 0.5496115

[[8]]
[1] 0.4842312

[[9]]
[1] 0.5412809

[[10]]
[1] 0.4489434
```