INTRODUCTION TO R

Bruin Actuarial Society

BAS.00
BRUIN ACTUARIAL SOCIETY

AGENDA

- Introduction
- Getting Started
- Data Manipulation with dplyr
- Plotting with ggplot2 package

BAS.00
BRUIN ACTUARIAL SOCIETY

INTRODUCTION

BAS.00
BRUIN ACTUARIAL SOCIETY

WHY R?

- R is a popular statistical programming language
  - Free
  - Great capabilities: from simple regression to machine learning
- Excel has its limitation:
  - Hard to implement statistical models
  - Graphing in Excel can also be challenging at times
- R can solve all the above problems and perform the task you can do in Excel faster

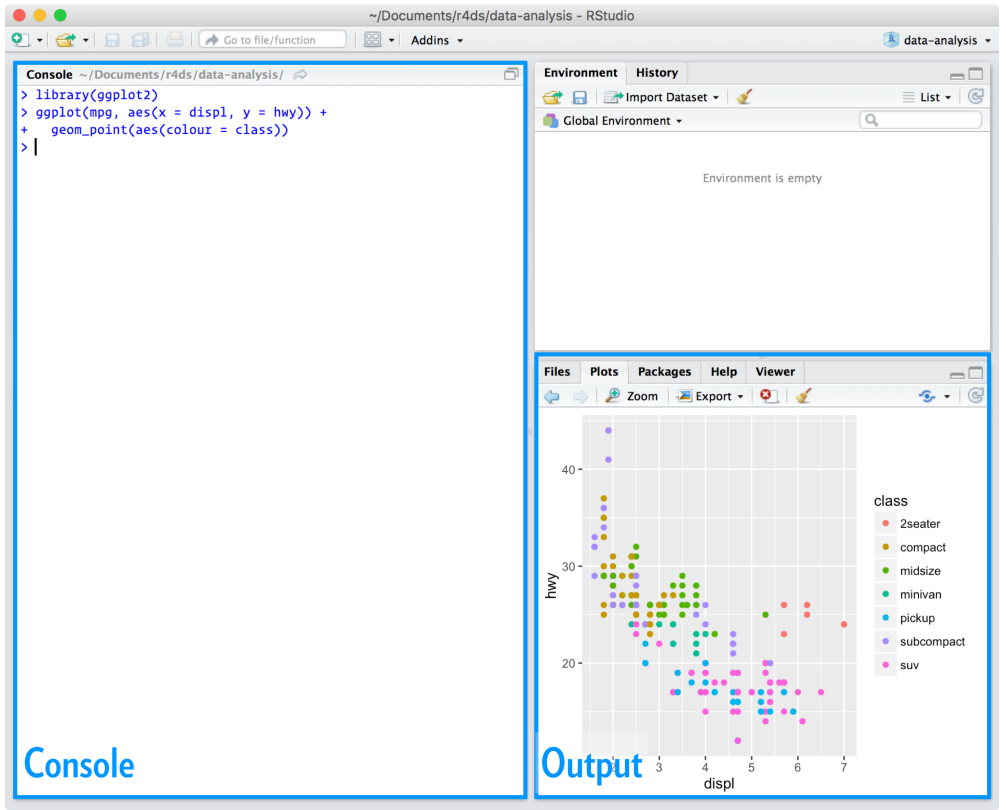
BAS.00
BRUIN ACTUARIAL SOCIETY

GETTING STARTED

BAS.00
BRUIN ACTUARIAL SOCIETY

STARTING R

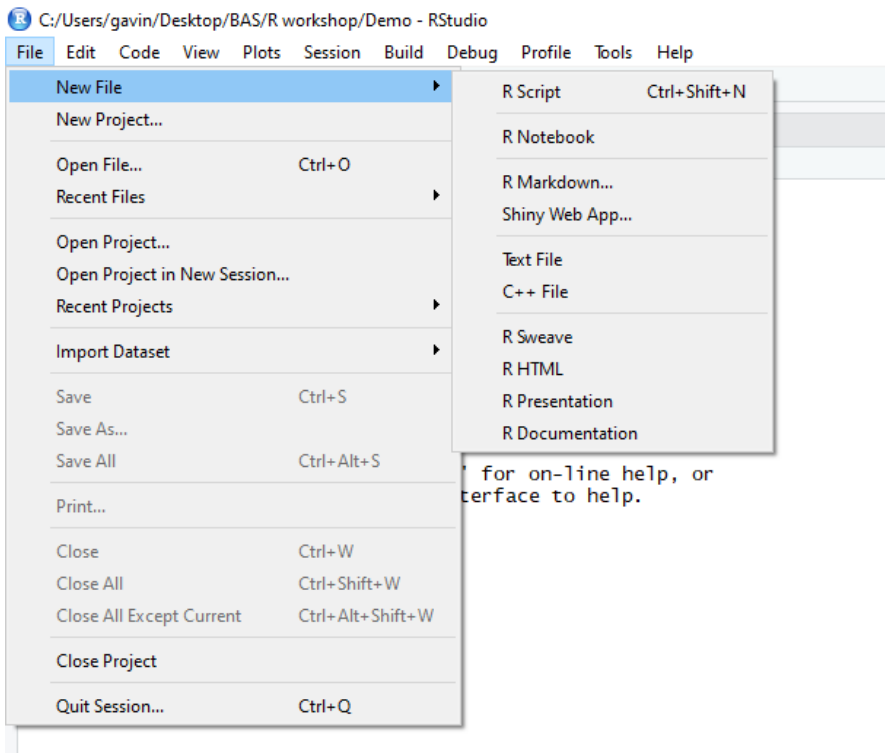
- We will use RStudio, an Integrated Development Environment, to program

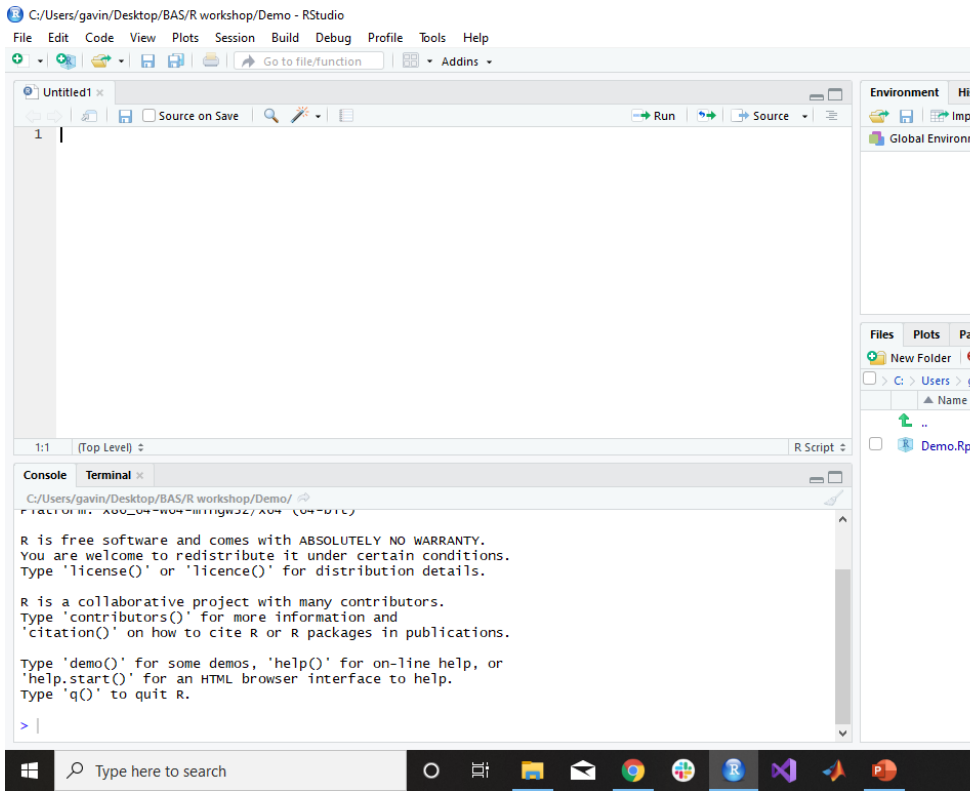


BAS.00
BRUIN ACTUARIAL SOCIETY

STARTING R

- To actually write and save your code, you will need a R script. We can add an R script under **“File”**, or **Ctrl + Shift + N (Window)/Cmd + Shift + N (Mac)**

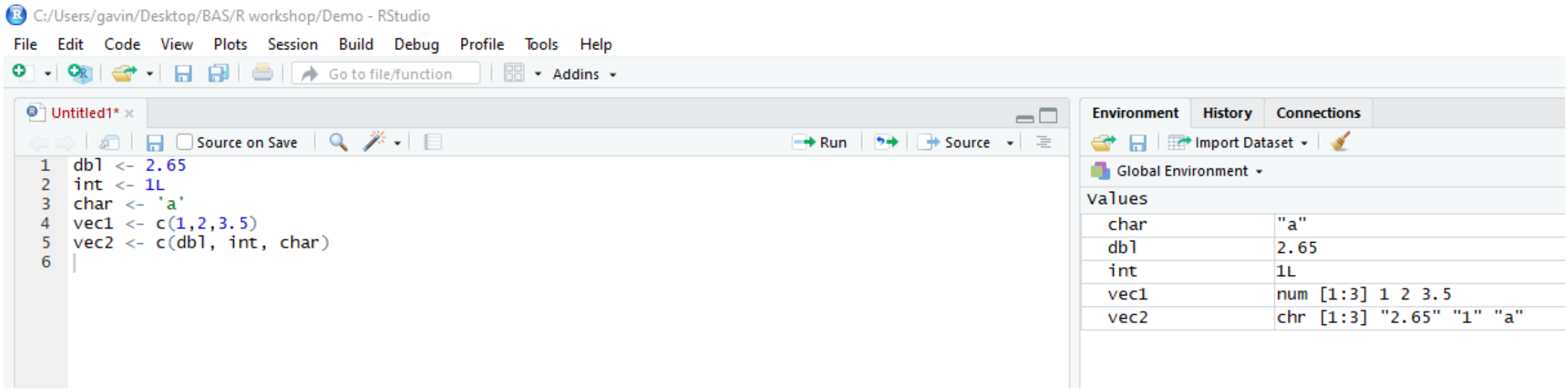




BAS.UCLA
BRUIN ACTUARIAL SOCIETY

ASSIGNMENT

- Use “**<-**” for assignment
- Use ”**c(…)**” to create vector



BAS.UCLA
BRUIN ACTUARIAL SOCIETY

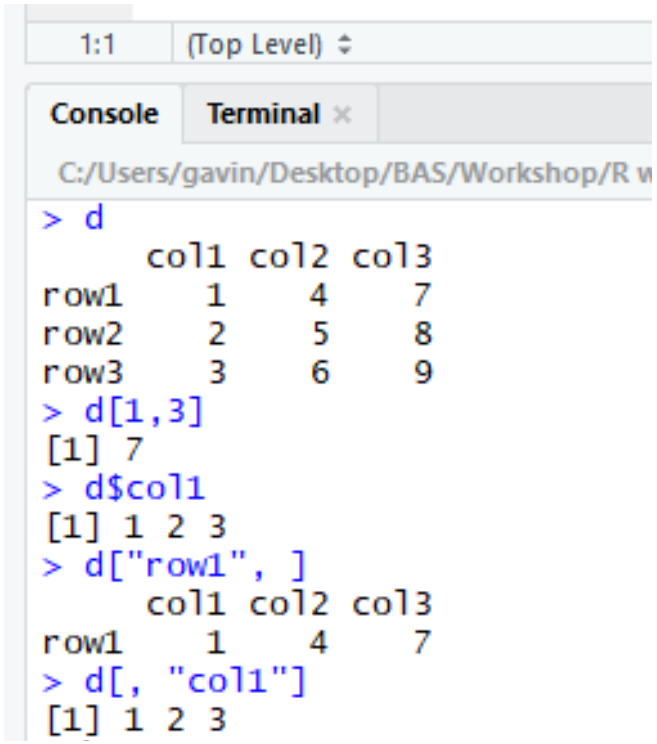
ARITHMETIC

- R does **element-wise operation** by default
- It also **recycle through** vector if their length does not match

```
> c(1,2,3)+c(4,5,6)
[1] 5 7 9
> c(1,2,3)*c(4,5,6)
[1]  4 10 18
> c(1,2,3,4) - c(5,6)
[1] -4 -4 -2 -2
>
```

SUBSETTING

- You can subset data table (matrix/dataframe/tibble) **using indices**
  - d[1, 3] represent 1st row 3rd column of d
- You can also subset **using column name** of data table with **"$"**
- You can also subset using row name and column name



BAS.UB
BRUIN ACTUARIAL SOCIETY

PRACTICE TIME!

BAS.00
BRUIN ACTUARIAL SOCIETY

EXERCISE

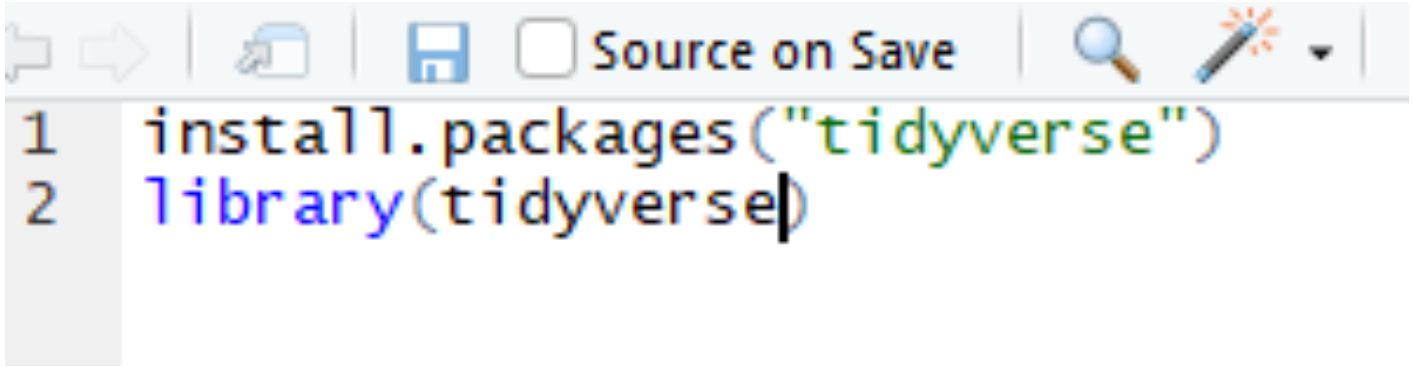
1. Create a vector from 1 to 10 and assign it the variable "v"
2. Increase the **odd** elements of v and decrease the **even** elements by 2 (Hint: vector cycling)
3. We will practice subsetting with the built-in dataset "iris"
   a. Type in iris on your console/script to examine the data
   b. Subset the 4th row and 2nd column of iris
   c. Subset the Petal Length column

BASIC DATA MANIPULATION WITH DPLYR

BAS.00
BRUIN ACTUARIAL SOCIETY

SETUP

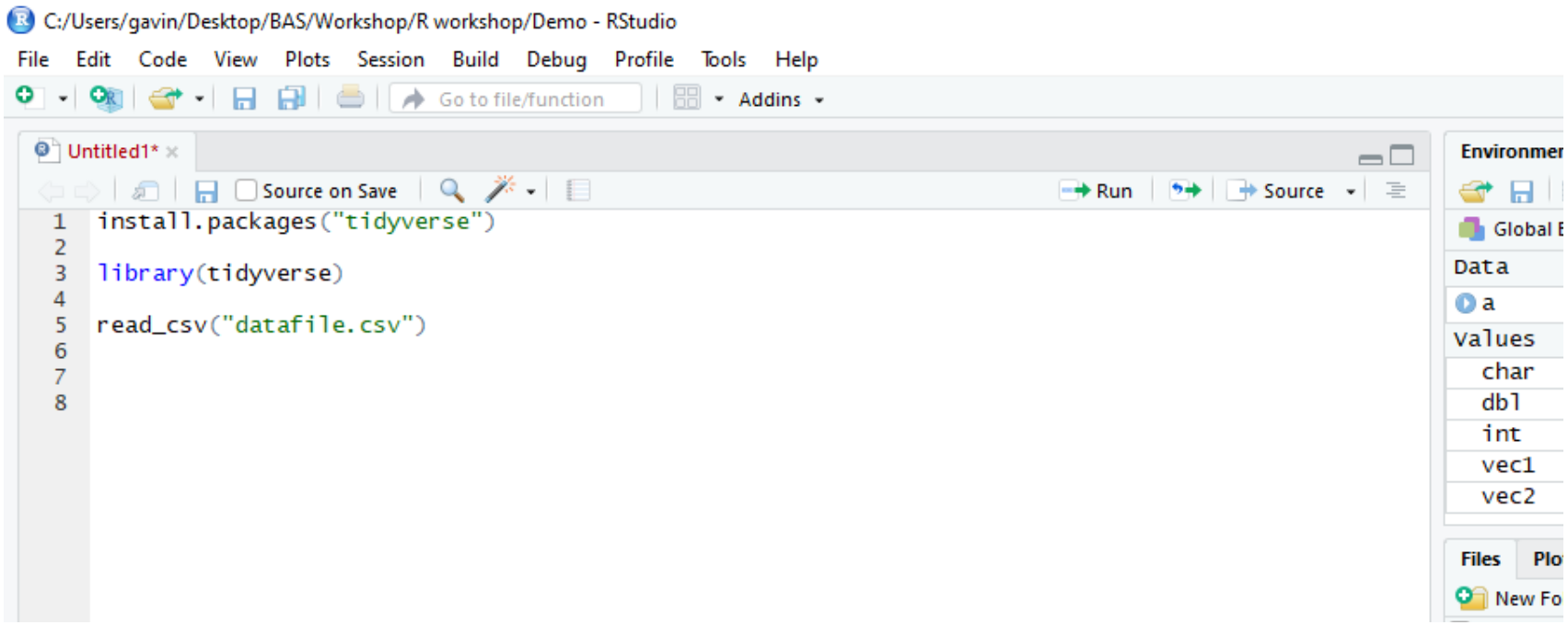
- We can install and load “packages” from internet in R
- We will load “tidyverse” package, which includes the “dplyr” package and “ggplot2” package



BAS.UD
BRUIN ACTUARIAL SOCIETY

READING DATA

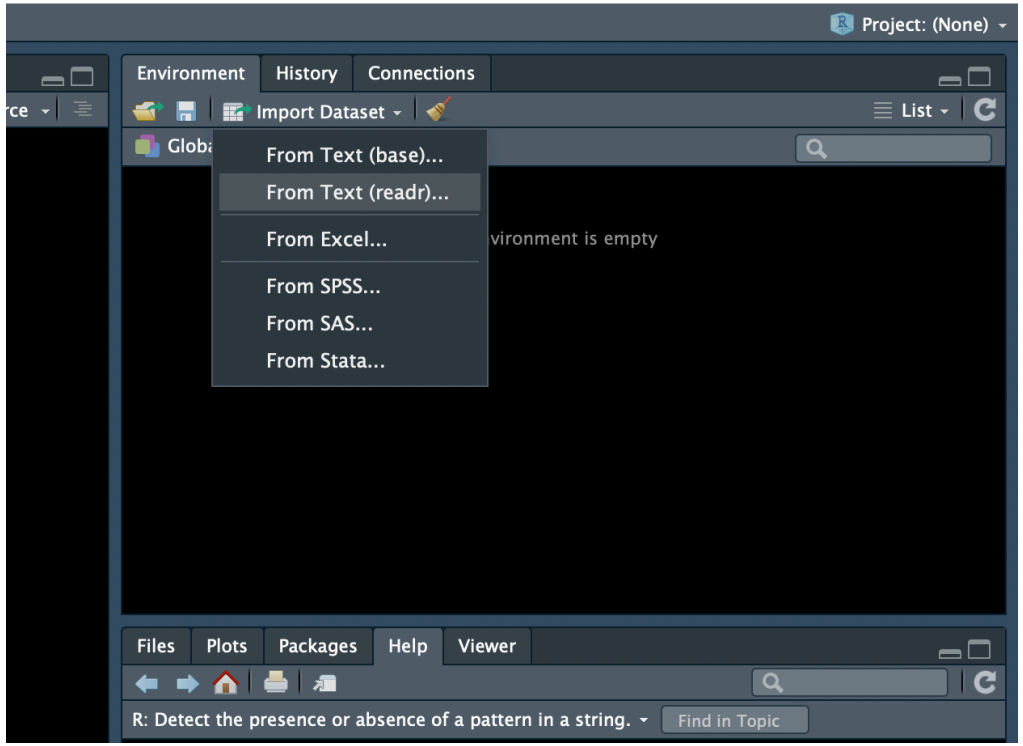
- R has build-in functions for reading data, but generally they are slow
- The “tidyverse” package contains the “readr” package that reads data quickly
- Usually, data are in “.csv” type

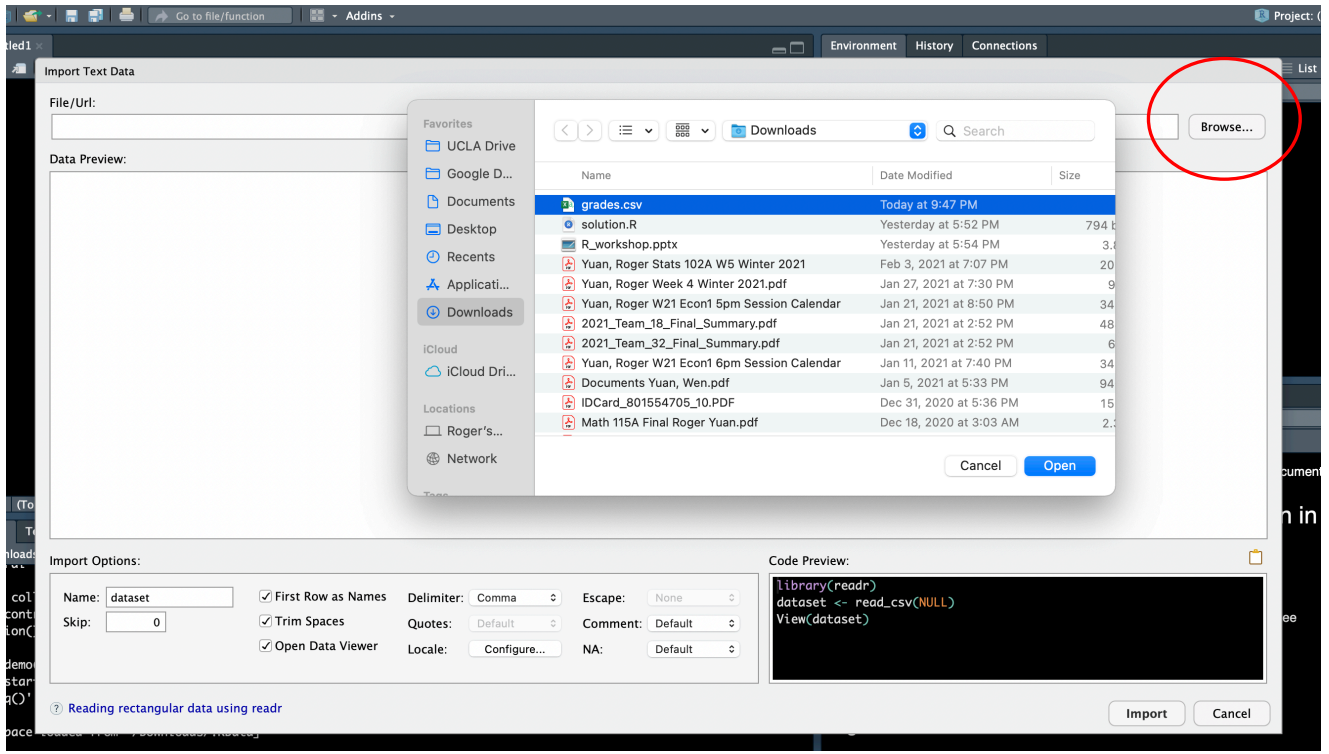


BAS.00
BRUIN ACTUARIAL SOCIETY

READING DATA SIMPLIFIED

- Importing CSV files through "Import Dataset" tab and browse to the desired file

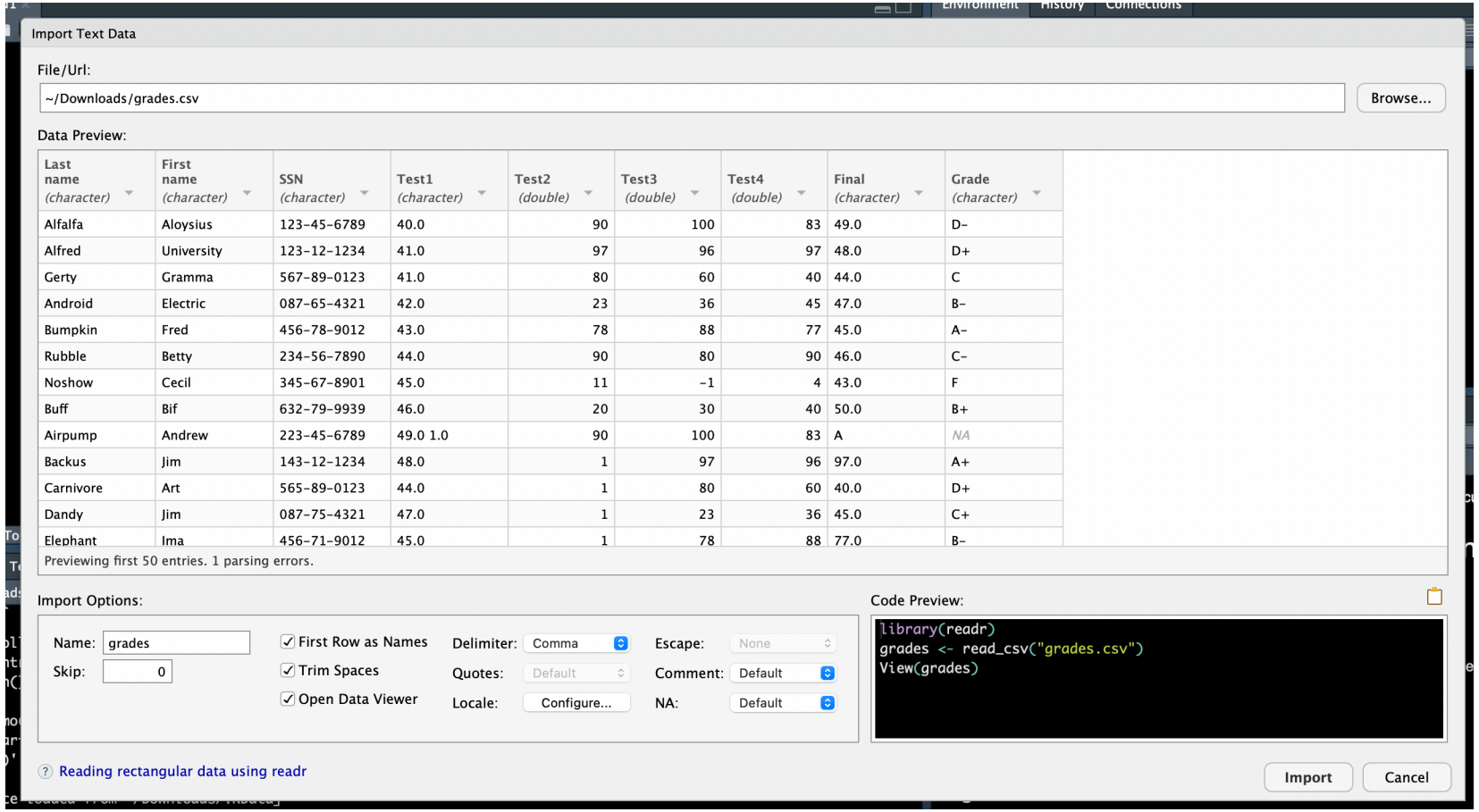




BAS.00
BRUIN ACTUARIAL SOCIETY

READING DATA SIMPLIFIED

- Click "Import" and the dataset will appear in the global environment



BAS.00
BRUIN ACTUARIAL SOCIETY

FILTER() FUNCTION

- The filter() function extracts rows by their value
- The syntax for filter is filter(data, logical_expression_for_extracting_rows)
- The filter function ignore rows that evaluates to **FALSE and NA**

Console Terminal

C:/Users/gavin/Desktop/BAS/Workshop/R workshop/Demo/

```
> filter(iris, Petal.Width > 0.3)
    Sepal.Length Sepal.Width Petal.Length Petal.Width    Species
1            5.4         3.9          1.7         0.4     setosa
2            5.7         4.4          1.5         0.4     setosa
3            5.4         3.9          1.3         0.4     setosa
4            5.1         3.7          1.5         0.4     setosa
5            5.1         3.3          1.7         0.5     setosa
6            5.0         3.4          1.6         0.4     setosa
7            5.4         3.4          1.5         0.4     setosa
8            5.0         3.5          1.6         0.6     setosa
9            5.1         3.8          1.9         0.4     setosa
10           7.0         3.2          4.7         1.4 versicolor
11           6.4         3.2          4.5         1.5 versicolor
12           6.9         3.1          4.9         1.5 versicolor
```

BAS.00
BRUIN ACTUARIAL SOCIETY

SELECT() FUNCTION

- The select() function extracts columns by their value
- The syntax for select() is select(data, col_names_condition)
- There are helpers function function for select():
  - start_with(): select name that start with whatever you specify
  - ends_with(): select name that ends with whatever you specify
  - contains(): select names that contain a specific string

SELECT() FUNCTION

Console Terminal ×

C:/Users/gavin/Desktop/BAS/Workshop/R workshop/Demo/

```
> select(iris, Petal.Width)
    Petal.Width
1           0.2
2           0.2
3           0.2
4           0.2
5           0.2
6           0.4
7           0.3
8           0.2
9           0.2
10          0.1
11          0.2
12          0.2
```

Console Terminal ×

C:/Users/gavin/Desktop/BAS/Workshop/R workshop/De

```
> select(iris, ends_with("Length"))
    Sepal.Length Petal.Length
1            5.1          1.4
2            4.9          1.4
3            4.7          1.3
4            4.6          1.5
5            5.0          1.4
6            5.4          1.7
7            4.6          1.4
8            5.0          1.5
9            4.4          1.4
10           4.9          1.5
11           5.4          1.5
12           4.8          1.6
13           4.8          1.4
```

BAS.00
BRUIN ACTUARIAL SOCIETY

MUTATE() FUNCTION

- Compute new variable from existing ones and add it to the data table
- Syntax: mutate(data, var_name = [operation w/ existing columns])

```
> mutate(iris, lpw_p = Petal.Length/Petal.Width)
    Sepal.Length Sepal.Width Petal.Length Petal.Width    Species     lpw_p
1            5.1         3.5          1.4         0.2     setosa  7.000000
2            4.9         3.0          1.4         0.2     setosa  7.000000
3            4.7         3.2          1.3         0.2     setosa  6.500000
4            4.6         3.1          1.5         0.2     setosa  7.500000
5            5.0         3.6          1.4         0.2     setosa  7.000000
6            5.4         3.9          1.7         0.4     setosa  4.250000
7            4.6         3.4          1.4         0.3     setosa  4.666667
8            5.0         3.4          1.5         0.2     setosa  7.500000
9            4.4         2.9          1.4         0.2     setosa  7.000000
10           4.9         3.1          1.5         0.1     setosa 15.000000
11           5.4         3.7          1.5         0.2     setosa  7.500000
12           4.8         3.4          1.6         0.2     setosa  8.000000
13           4.8         3.0          1.4         0.1     setosa 14.000000
14           4.3         3.0          1.1         0.1     setosa 11.000000
15           5.8         4.0          1.2         0.2     setosa  6.000000
16           5.7         4.4          1.5         0.4     setosa  3.750000
17           5.4         3.9          1.3         0.4     setosa  3.250000
18           5.1         3.5          1.4         0.3     setosa  4.666667
19           5.7         3.8          1.7         0.3     setosa  5.666667
```

BAS.00
BRUIN ACTUARIAL SOCIETY

PRACTICE TIME!

BAS.00
BRUIN ACTUARIAL SOCIETY

EXERCISE

1. Using the filter() function, capture all the rows that has sepal width greater than 1 in the dataset iris
2. Using the select() function and its helper functions, capture all the columns that start with the letter “s”
3. Create a separate column in iris that contains the sum of the sepal width and length

PLOTTING WITH GGPLOT2

BAS.00
BRUIN ACTUARIAL SOCIETY

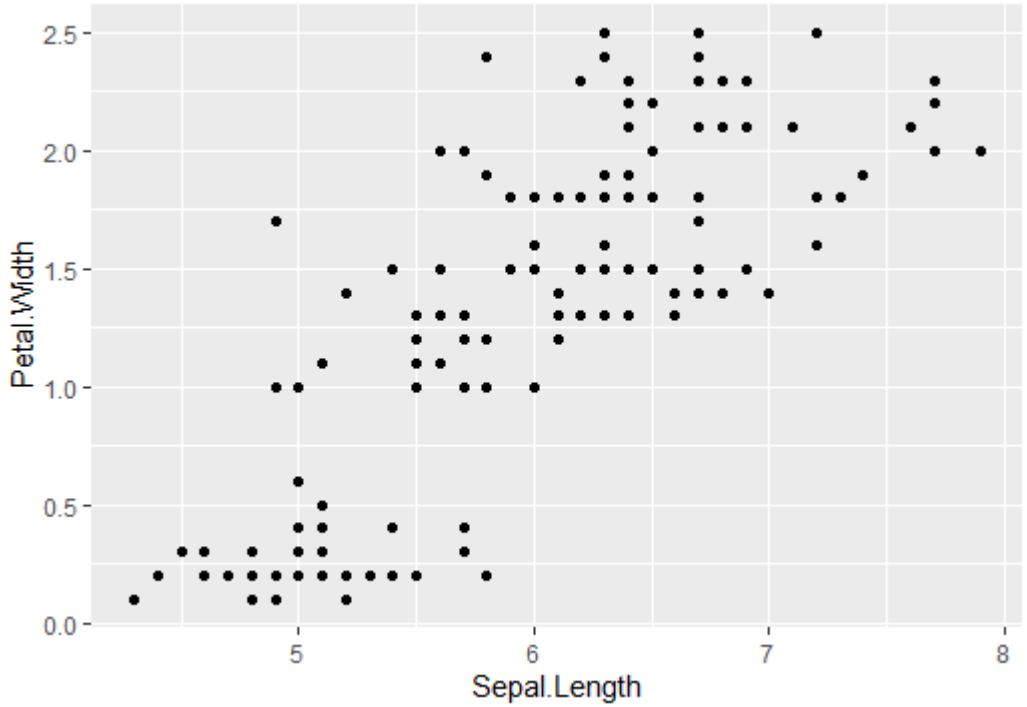
WHAT IS GGPLOT2?

- ggplot2 is a R package that is very useful for data visualization
- "gg" stands for grammars of graphics
- It provides a systematic way to create visually-pleasing graphs
- Basic syntax
  - ggplot(data = data_name, aes(x=..., y=…)) + geom_...

BAS.UD
BRUIN ACTUARIAL SOCIETY

SCATTERPLOT

- To create a scatterplot, we use the **geom_point()** function
  - ggplot(data = data_name, aes(x=..., y=...)) + geom_point()



Untitled1* | title

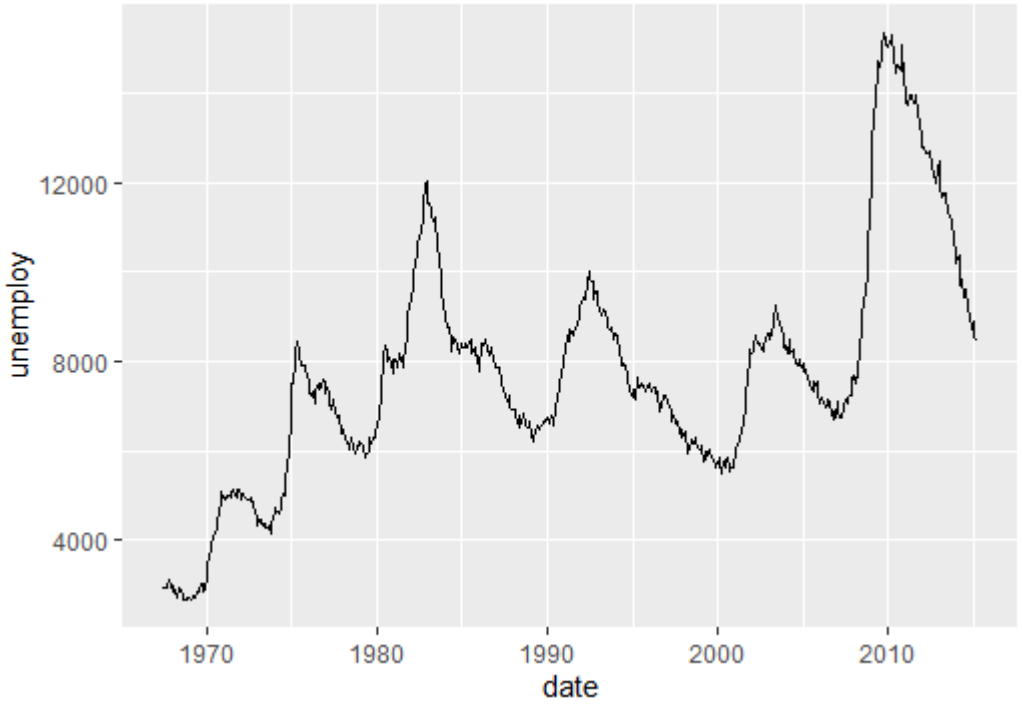
Source on Save

```
1  install.packages("tidyverse")
2
3  library(tidyverse)
4
5  ##we will use the build-in dataset "iris"
6
7  plot1 <- ggplot(data = iris,
8                  aes(x = Sepal.Length, y = Petal.Width))+
9           geom_point()
10 plot1
11
```

BAS.00
BRUIN ACTUARIAL SOCIETY

LINE PLOT

- To create a line plot, we use the **geom_line()** function
  - ggplot(data = data_name, aes(x=..., y=...)) + geom_line()



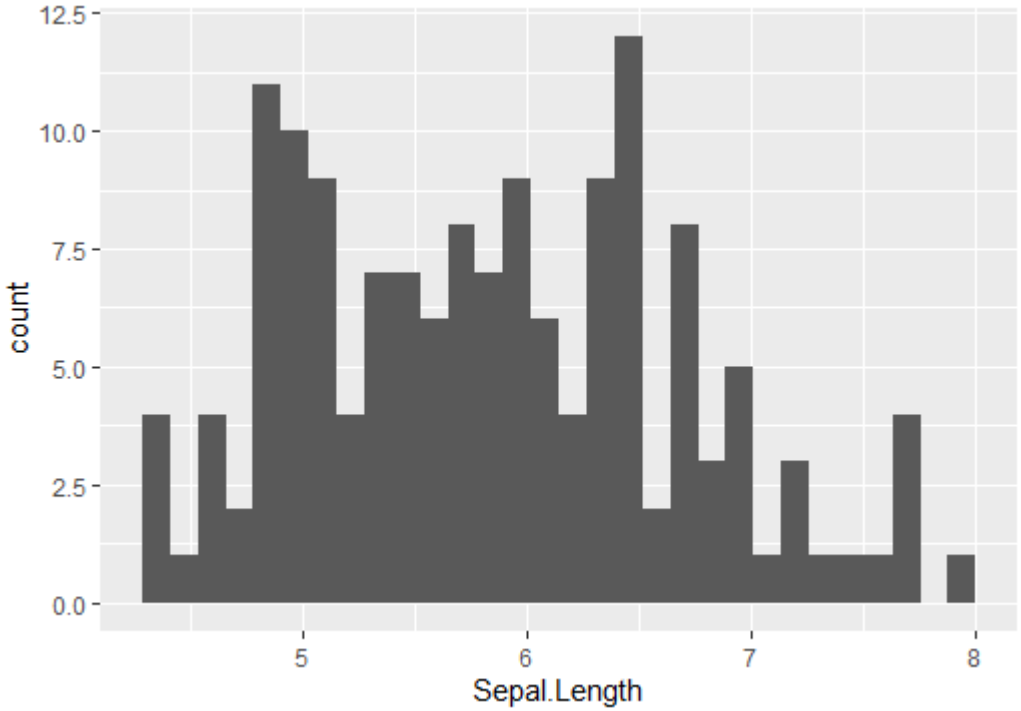

```
library(tidyverse)

plot2 <- ggplot(data = economics,
                aes(date, unemploy))+
  geom_line()+
plot2
```

BAS.00
BRUIN ACTUARIAL SOCIETY

HISTOGRAM

- To create a histogram, we use the **geom_histogram()** function
  - ggplot(data = data_name, aes(x=..., y=...)) + geom_histogram()




```
3  library(tidyverse)
4
5  ##we will use the build-in dataset "iris"
6
7  plot3 <- ggplot(data = iris,
8                  aes(Sepal.Length))+
9           geom_histogram()
10 plot3
11
```

BAS.00
BRUIN ACTUARIAL SOCIETY

PRACTICE TIME!

BAS.00
BRUIN ACTUARIAL SOCIETY

EXERCISE

1. Type on mtcars on your script or console and take a look at the data. We will be using this dataset
2. Create a scatterplot for “disp” and “mpg” and save it as splot
3. Create a line graph for “wt” and “qsec” and save it as line_graph
4. Create histogram for “cyl” and save it as histo

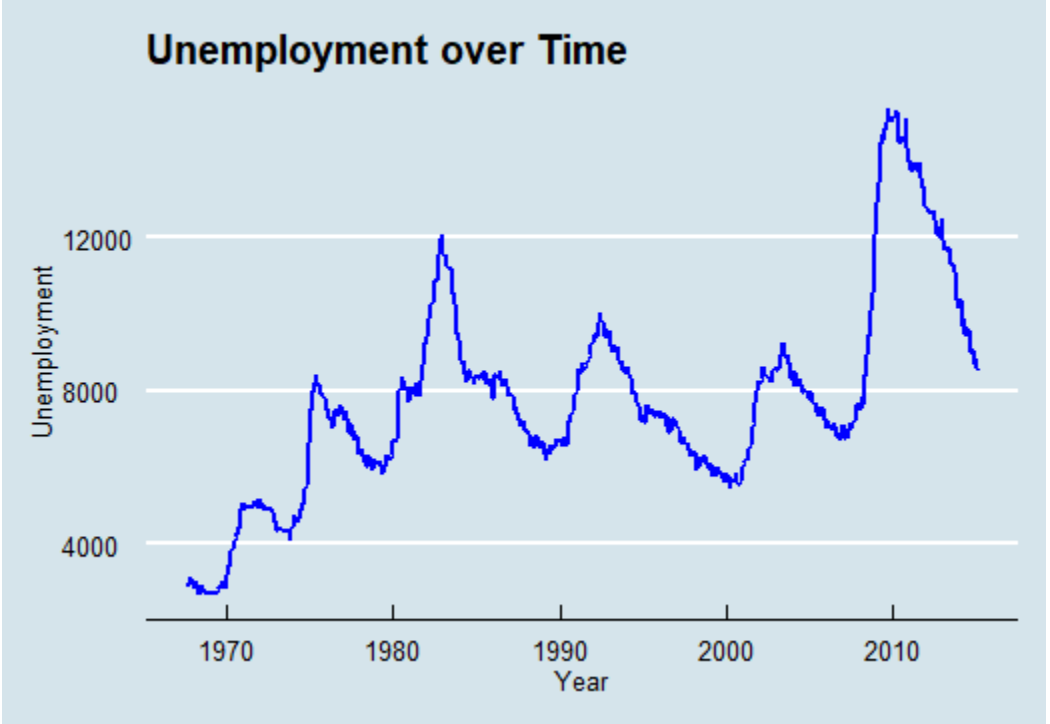
HOW TO MAKE PLOT PRETTIER

- We can specify additional elements in the ggplot to make plots look nicer!
- Example
  - Rename axis
  - Add a title
  - Color
  - Themes
  - Add a smoother function
  - And more!

BAS.DD
BRUIN ACTUARIAL SOCIETY

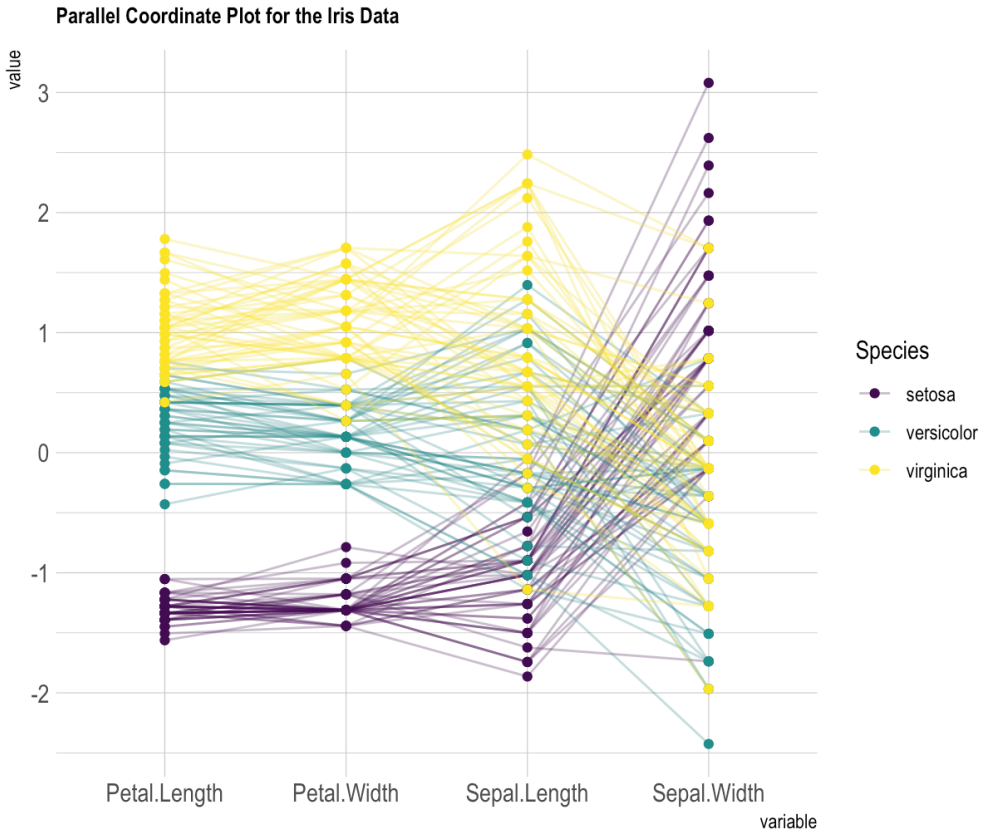
LINE PLOT REVISITED

```
3  library(tidyverse)
4
5  install.packages("ggthemes")
6  library(ggthemes)
7  plot2_nice <- ggplot(data = economics,
8                       aes(date, unemploy))+
9            geom_line(size = 1, color = "blue")+
10           xlab("Year")+
11           ylab("Unemployment")+
12           ggtitle("Unemployment over Time")+
13           theme_economist()
14 plot2_nice
15
```



BAS.00
BRUIN ACTUARIAL SOCIETY

GGPLOT2 IS VERY POWERFUL



BAS.00
BRUIN ACTUARIAL SOCIETY

Question?