



# Introduction to R for Plant Pathologists

## APS Online Webinar

Sydney Everhart<sup>1</sup>, Nikita Gambhir<sup>1</sup>, and Katie Gold<sup>2</sup>

<sup>1</sup>University of Nebraska-Lincoln

<sup>2</sup>Cornell University

February 19<sup>th</sup> and 26<sup>th</sup>, 2020

# Workshop Structure

- February 19th:
  - Part 1. Introduction to R (Sydney Everhart)
  - Online recording to be posted at: [apsnet.org/edcenter/resources/Webinars](https://apsnet.org/edcenter/resources/Webinars)
- February 26th:
  - Part 2a. Data Manipulation and Analysis (Nikita Gambhir)
  - Part 2b. Data Visualization (Katie Gold)
  - Part 2c. Starting a Project (Katie Gold)



# Part 2a. Data Manipulation and Analysis

Nikita Gambhir

# Part 2a: Data Manipulation and Analysis

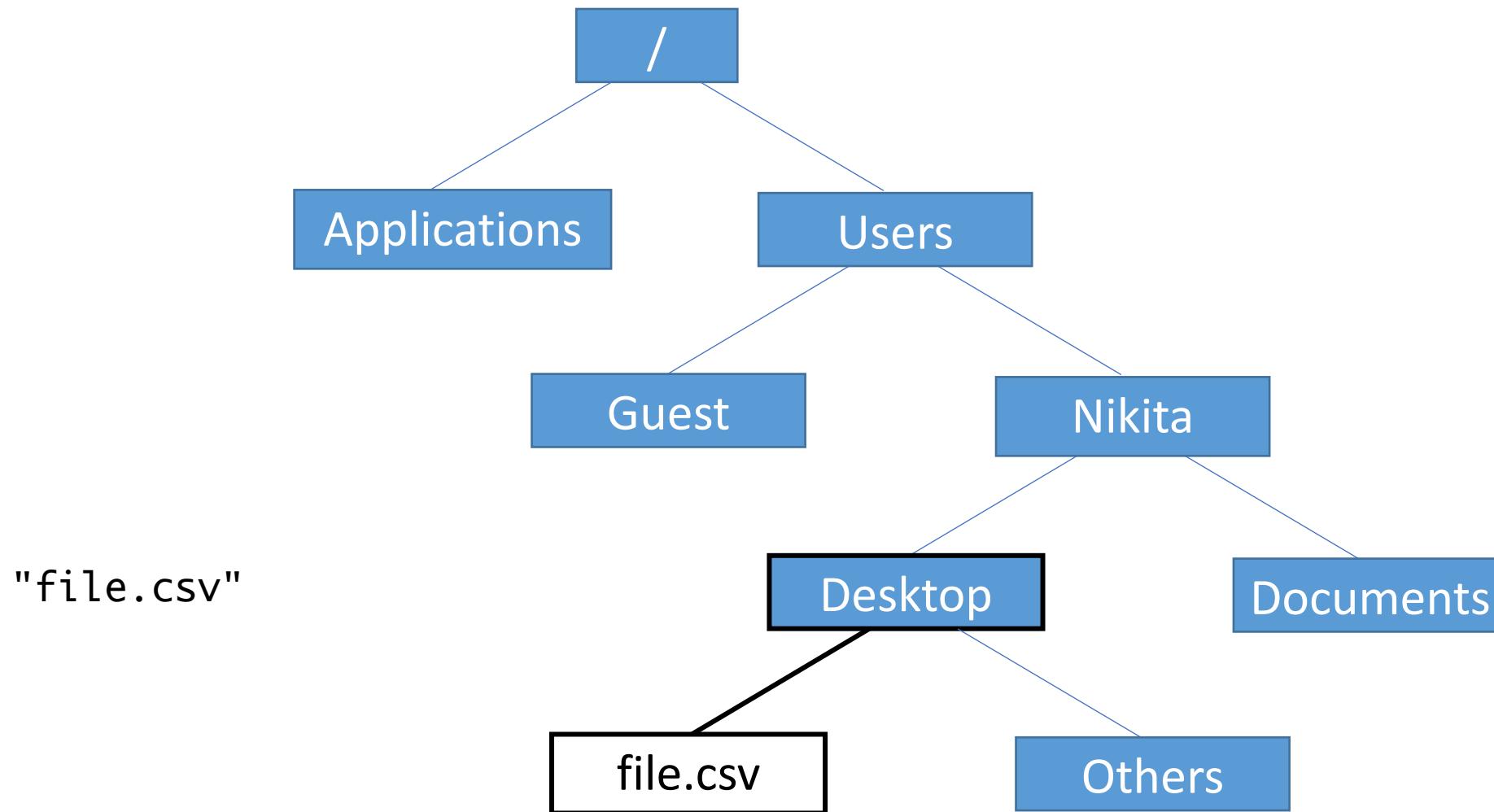
- Import data
- File formats
- Packages
- Analysis
- *Coding at the command line!*

	Treatment	Yield_bu_per_acre	Severity
1	Control	173.82	5.5
2	Control	174.23	5.6
3	Control	173.57	5.4
4	Control	173.61	5.4
5	Control	174.19	5.6
6	Control	173.8	5.5
7	Fungicide_A	173.98	5.1
8	Fungicide_A	174.27	5.2
9	Fungicide_A	173.61	5
10	Fungicide_A	173.88	5
11	Fungicide_A	174.17	5.2
12	Fungicide_A	173.49	5.1
13	Fungicide_B	175.98	4.1
14	Fungicide_B	175.58	3.9
15	Fungicide_B	175.75	4.2
16	Fungicide_B	175.88	4
17	Fungicide_B	175.68	4.1
18	Fungicide_B	175.95	4.2

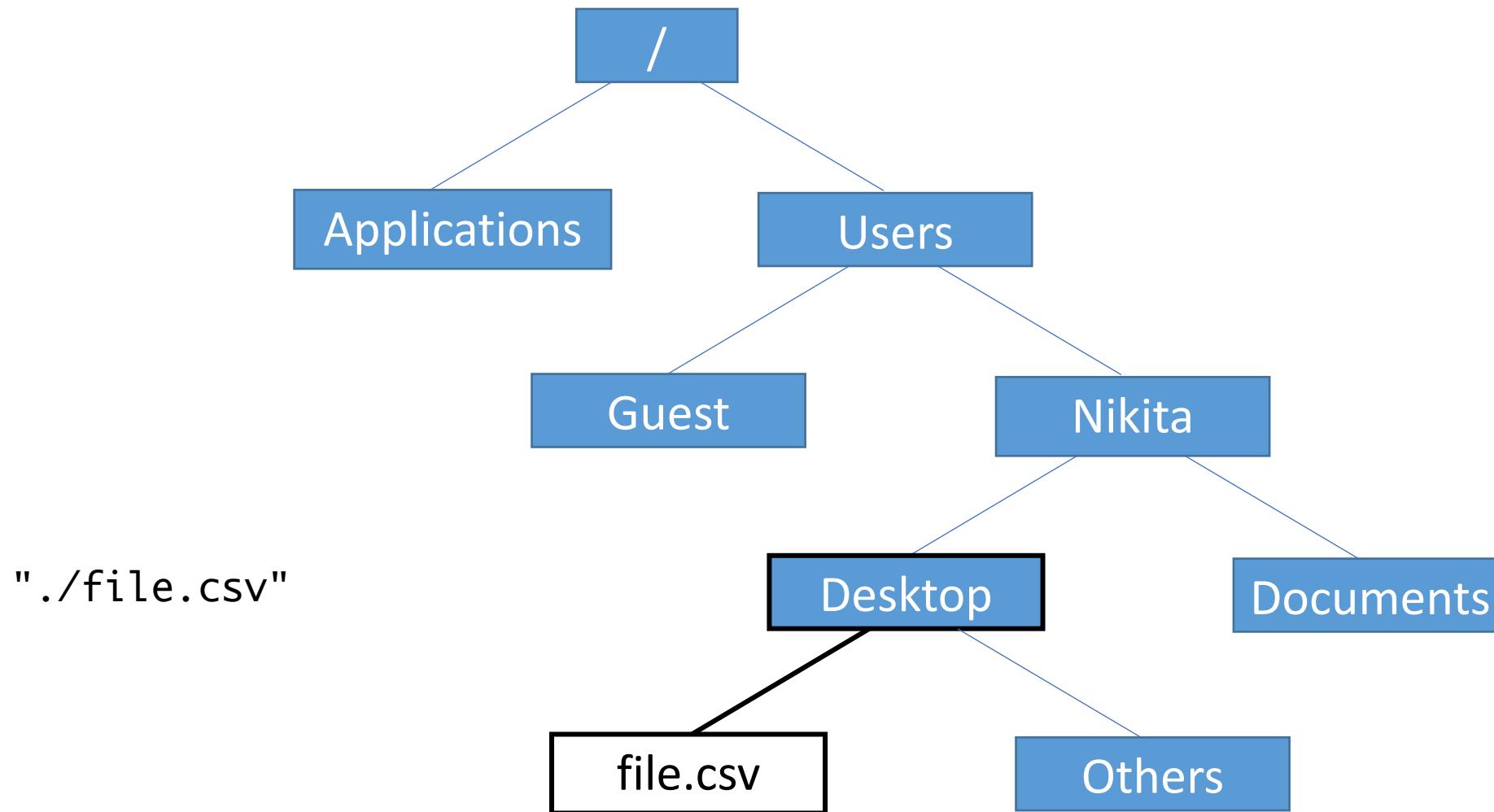


# How to specify the path of a file?

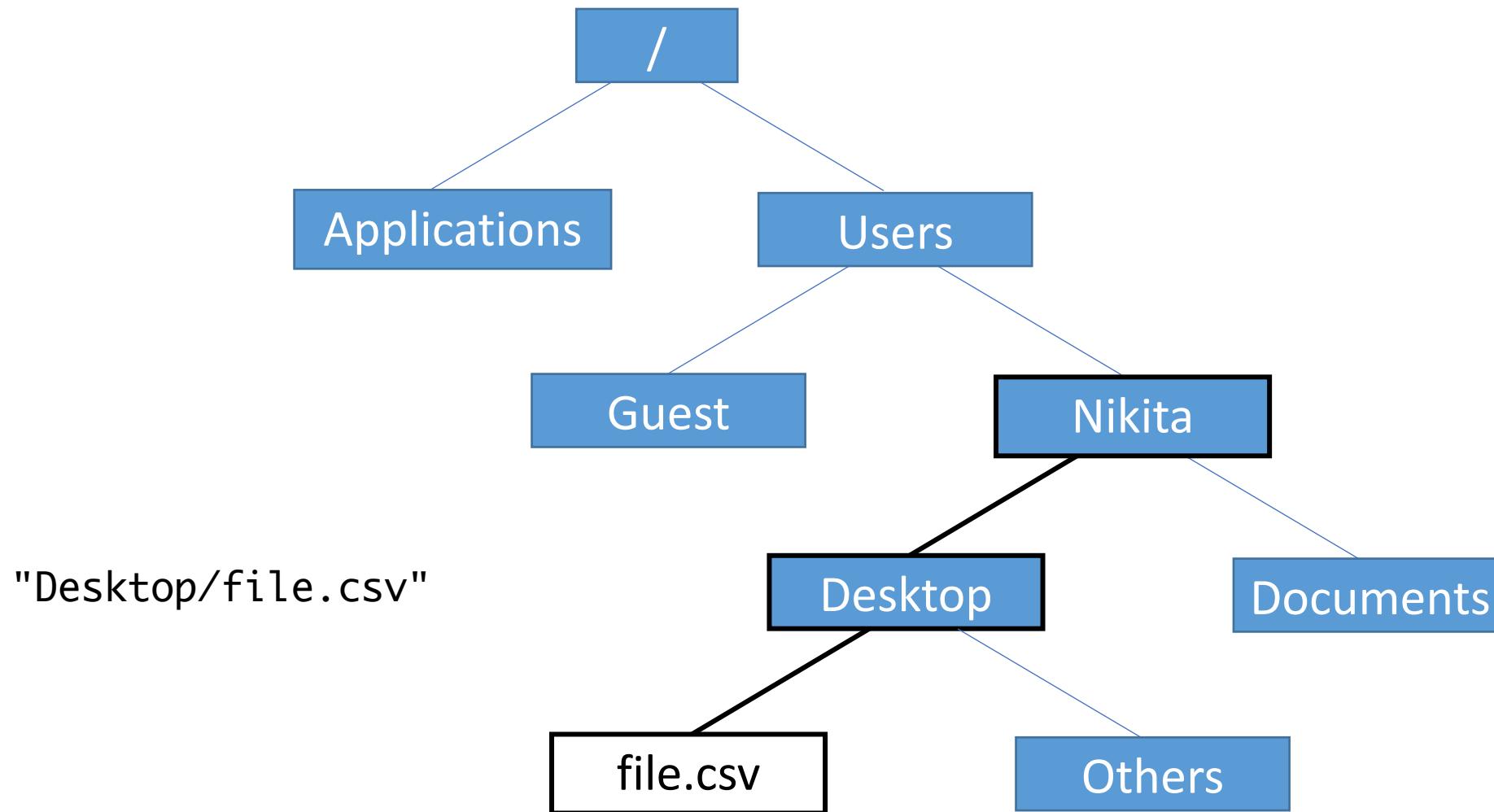
# Directory Structure



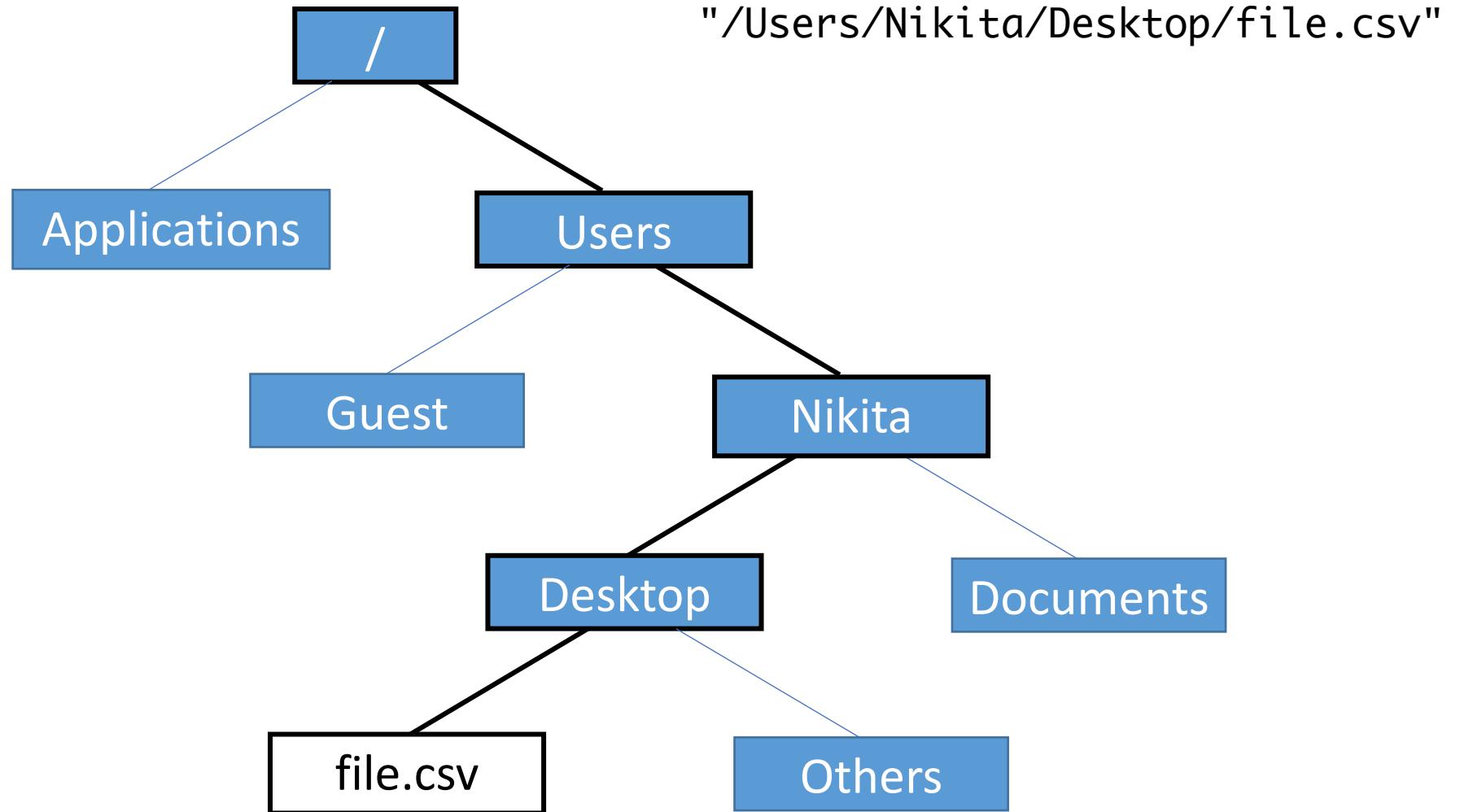
# Directory Structure



# Directory Structure



# Directory Structure



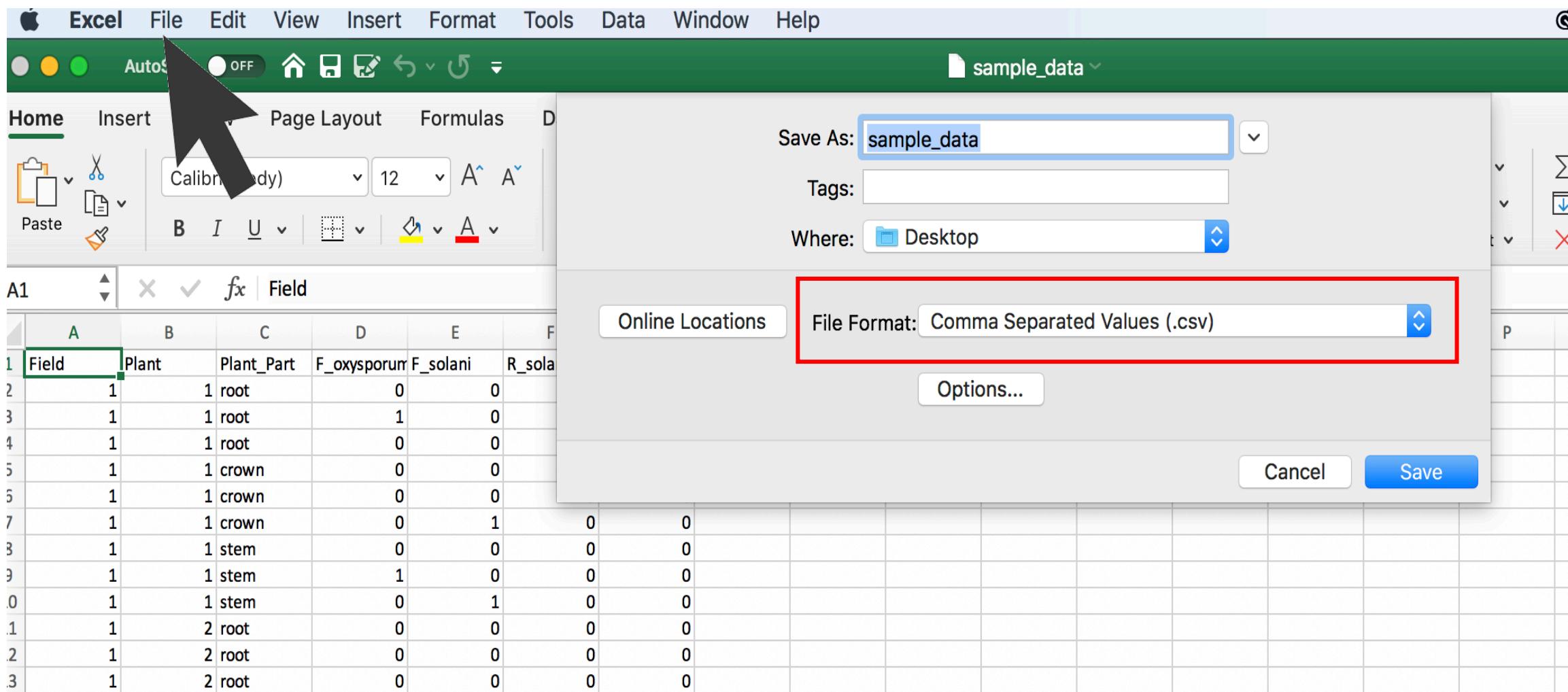


# Data file formats

# Data file formats

XLS	CSV
Binary file format for all eXceL Sheets	Plain text format with <b>Comma Separated Values</b>
Can be opened with Microsoft Excel only	Can be opened with any text editor (like notepad), MS Excel, etc.
Can do operations on data like formatting, formulas, macros, etc.	Can only store data
Consumes more memory	Consumes less memory
Difficult to manipulate in programming languages	Easy to manipulate

# How to save a data file in '.csv' format?





# Popular data science packages



## R packages for data science

The tidyverse is an opinionated **collection of R packages** designed for data science. All packages share an underlying design philosophy, grammar, and data structures.

Install the complete tidyverse with:

```
install.packages("tidyverse")
```

# Tidy Data

Treatment	Yield	Severity
Control	173.82	5.5
Control	174.23	5.6
Control	173.57	5.4
Control	173.61	5.4
Control	174.19	5.6
Control	173.8	5.5
Fungicide_A	173.98	5.1
Fungicide_A	174.27	5.2
Fungicide_A	173.61	5
Fungicide_A	173.88	5
Fungicide_A	174.17	5.2
Fungicide_A	173.49	5.1

# Tidy Data

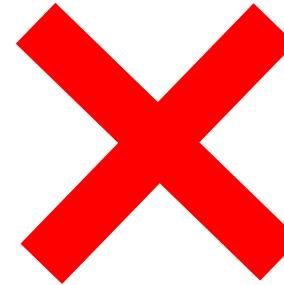
Treatment	Yield	Severity
Control	173.82	5.5
Control	174.23	5.6
Control	173.57	5.4
Control	173.61	5.4
Control	174.19	5.6
Control	173.8	5.5
Fungicide_A	173.98	5.1
Fungicide_A	174.27	5.2
Fungicide_A	173.61	5.1
Fungicide_A	173.88	5.1
Fungicide_A	174.17	5.2
Fungicide_A	173.49	5.1

variables

# Tidy Data

Treatment	Yield	Severity
Control	173.82	5.5
Control	174.23	5.6
Control	173.57	5.4
Control	173.61	5.4
Control	174.19	5.6
Control	173.8	5.5
Fungicide_A	173.98	5.1
Fungicide_A	174.27	5.2
Fungicide_A	173.61	5.1
Fungicide_A	173.88	5.1
Fungicide_A	174.17	5.2
Fungicide_A	173.49	5.1

variables



Treatment	Control	Control	Control	Control
Yield	173.82	174.23	173.57	173.61
Severity	5.5	5.6	5.4	5.4

# Tidy Data

Treatment	Yield	Severity
Control	173.82	5.5
Control	174.23	5.6
Control	173.57	5.4
Control	173.61	5.4
Control	174.19	5.6
Control	173.8	5.5
Fungicide_A	173.98	5.1
Fungicide_A	174.27	5.2
Fungicide_A	173.61	5.0
Fungicide_A	173.88	5.0
Fungicide_A	174.17	5.2
Fungicide_A	173.49	5.1

variables

Treatment	Yield	Severity
Control	173.82	5.5
Control	174.23	5.6
Control	173.57	5.4
Control	173.61	5.4
Control	174.19	5.6
Control	173.8	5.5
Fungicide_A	173.98	5.1
Fungicide_A	174.27	5.2
Fungicide_A	173.61	5.0
Fungicide_A	173.88	5.0
Fungicide_A	174.17	5.2
Fungicide_A	173.49	5.1

observations

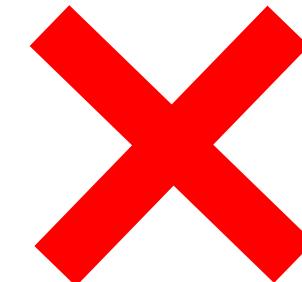
# Tidy Data

Treatment	Yield	Severity
Control	173.82	5.5
Control	174.23	5.6
Control	173.57	5.4
Control	173.61	5.4
Control	174.19	5.6
Control	173.8	5.5
Fungicide_A	173.98	5.1
Fungicide_A	174.27	5.2
Fungicide_A	173.61	5.0
Fungicide_A	173.88	5.2
Fungicide_A	174.17	5.2
Fungicide_A	173.49	5.1

variables

Treatment	Yield	Severity
Control	173.82	5.5
Control	174.23	5.6
Control	173.57	5.4
Control	173.61	5.4
Control	174.19	5.6
Control	173.8	5.5
Fungicide_A	173.98	5.1
Fungicide_A	174.27	5.2
Fungicide_A	173.61	5.0
Fungicide_A	173.88	5.2
Fungicide_A	174.17	5.2
Fungicide_A	173.49	5.1

observations



Treatment	Observed_Value
Control	173.82
	5.5
Control	174.23
	5.6
Control	173.57
	5.4
Fungicide_A	173.98
Fungicide_A	5.1

1<sup>st</sup>  
Observation

# Tidy Data

Treatment	Yield	Severity
Control	173.82	5.5
Control	174.23	5.6
Control	173.57	5.4
Control	173.61	5.4
Control	174.19	5.6
Control	173.8	5.5
Fungicide_A	173.98	5.1
Fungicide_A	174.27	5.2
Fungicide_A	173.61	5.0
Fungicide_A	173.88	5.0
Fungicide_A	174.17	5.2
Fungicide_A	173.49	5.1

variables

Treatment	Yield	Severity
Control	173.82	5.5
Control	174.23	5.6
Control	173.57	5.4
Control	173.61	5.4
Control	174.19	5.6
Control	173.8	5.5
Fungicide_A	173.98	5.1
Fungicide_A	174.27	5.2
Fungicide_A	173.61	5.0
Fungicide_A	173.88	5.0
Fungicide_A	174.17	5.2
Fungicide_A	173.49	5.1

observations

Treatment	Yield	Severity
Control	173.82	5.5
Control	174.23	5.6
Control	173.57	5.4
Control	173.61	5.4
Control	174.19	5.6
Control	173.8	5.5
Fungicide_A	173.98	5.1
Fungicide_A	174.27	5.2
Fungicide_A	173.61	5.0
Fungicide_A	173.88	5.0
Fungicide_A	174.17	5.2
Fungicide_A	173.49	5.1

values

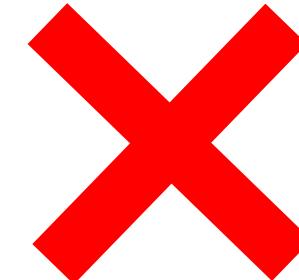
# Tidy Data

Treatment	Yield	Severity
Control	173.82	5.5
Control	174.23	5.6
Control	173.57	5.4
Control	173.61	5.4
Control	174.19	5.6
Control	173.8	5.5
Fungicide_A	173.98	5.1
Fungicide_A	174.27	5.2
Fungicide_A	173.61	5.0
Fungicide_A	173.88	5.0
Fungicide_A	174.17	5.2
Fungicide_A	173.49	5.1

variables

Treatment	Yield	Severity
Control	173.82	5.5
Control	174.23	5.6
Control	173.57	5.4
Control	173.61	5.4
Control	174.19	5.6
Control	173.8	5.5
Fungicide_A	173.98	5.1
Fungicide_A	174.27	5.2
Fungicide_A	173.61	5.0
Fungicide_A	173.88	5.0
Fungicide_A	174.17	5.2
Fungicide_A	173.49	5.1

observations



Treatment	Yield	Severity
Control_1	173.82	5.5
Control_2	174.23	5.6
Control_3	173.57	5.4
Control_4	173.61	5.4
Control_5	174.19	5.6
Control_6	173.8	5.5

# Tidy Data

Treatment	Yield	Severity
Control	173.82	5.5
Control	174.23	5.6
Control	173.57	5.4
Control	173.61	5.4
Control	174.19	5.6
Control	173.8	5.5
Fungicide_A	173.98	5.1
Fungicide_A	174.27	5.2
Fungicide_A	173.61	5.0
Fungicide_A	173.88	5.0
Fungicide_A	174.17	5.2
Fungicide_A	173.49	5.1

variables

Treatment	Yield	Severity
Control	173.82	5.5
Control	174.23	5.6
Control	173.57	5.4
Control	173.61	5.4
Control	174.19	5.6
Control	173.8	5.5
Fungicide_A	173.98	5.1
Fungicide_A	174.27	5.2
Fungicide_A	173.61	5.0
Fungicide_A	173.88	5.0
Fungicide_A	174.17	5.2
Fungicide_A	173.49	5.1

observations

Treatment	Yield	Severity
Control	173.82	5.5
Control	174.23	5.6
Control	173.57	5.4
Control	173.61	5.4
Control	174.19	5.6
Control	173.8	5.5
Fungicide_A	173.98	5.1
Fungicide_A	174.27	5.2
Fungicide_A	173.61	5.0
Fungicide_A	173.88	5.0
Fungicide_A	174.17	5.2
Fungicide_A	173.49	5.1

values



# Data Analysis

# Summary table for one-way ANOVA

	Df	Sum Sq	Mean Sq	F value	P- value	
Treatment	2	14.723	7.361	111.3	1.01e-09	***
Residuals	15	0.992	0.066			

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1



# Let's Get Coding!

# Summary of Part 2a.

- `mutate()` – adds new column to data frame
- `select()` - keeps only chosen variables
- `summarize()` – summarizes multiple values to a single value
- `group_by()` – groups data by chosen variables so that further operations can be performed
- `filter()` – finds rows of a variable that meet a certain condition



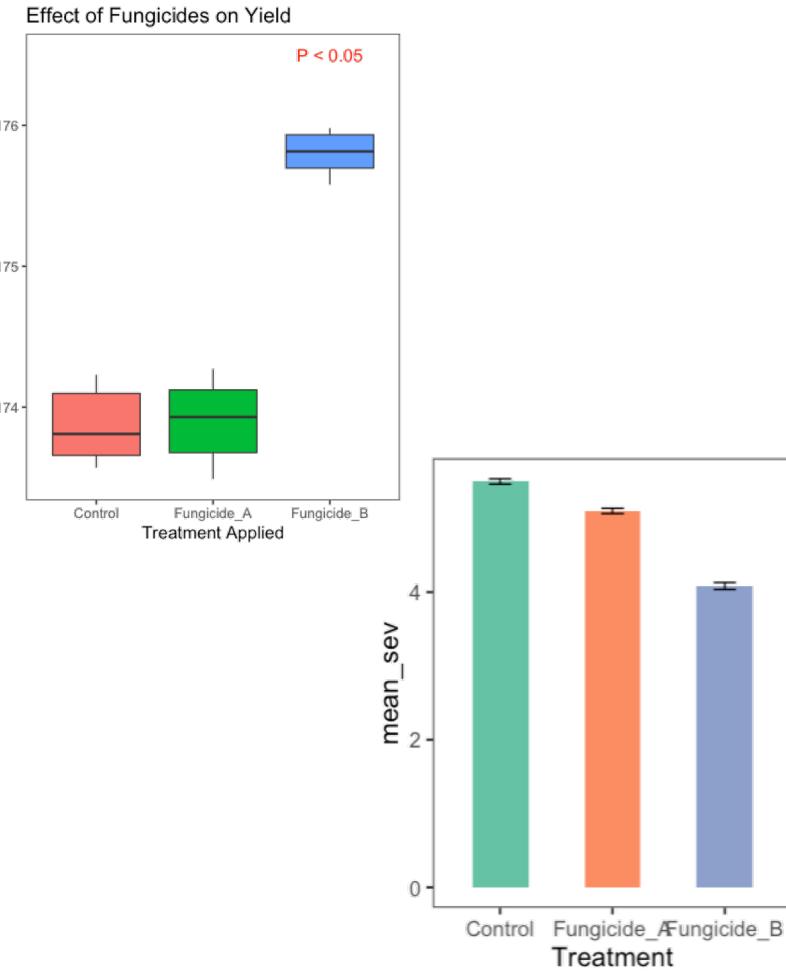
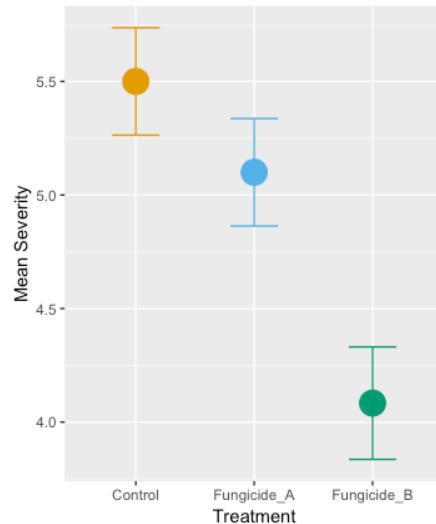
# Part 2b. Data Visualization

Katie Gold

# Part 2b: Visualization

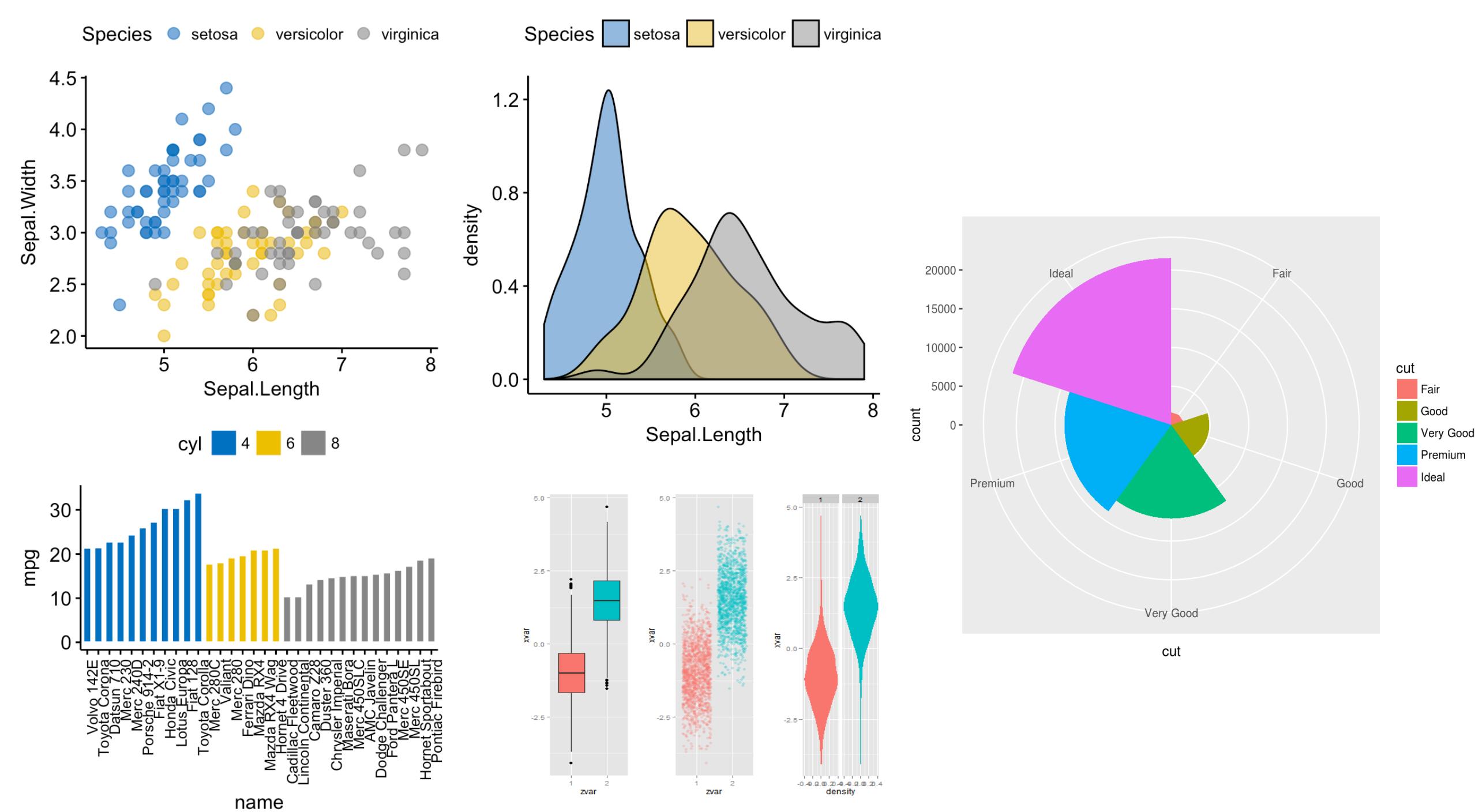
Use ggplot2 to:

- Create different types of plots
- Save plots





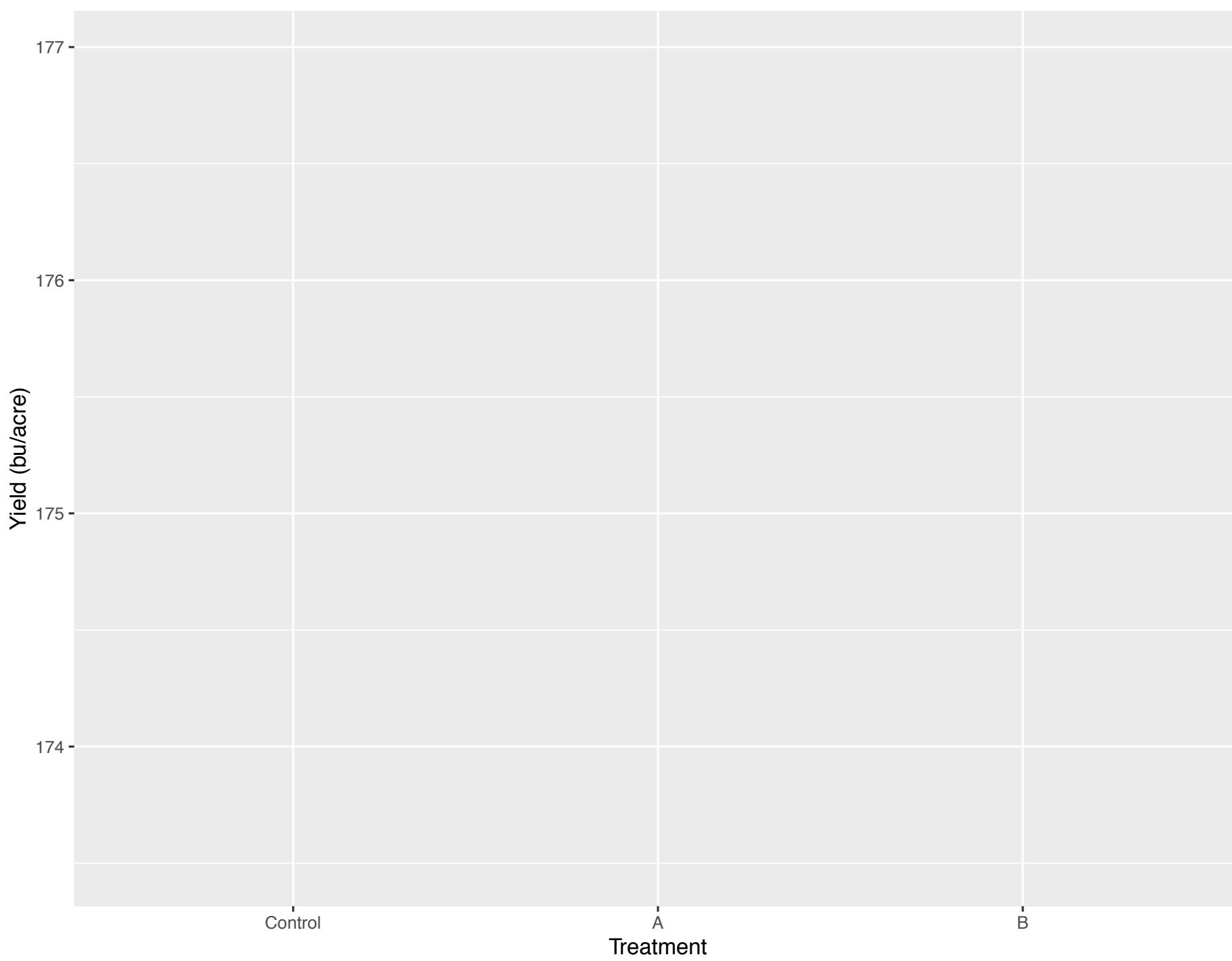
# What is ggplot2?

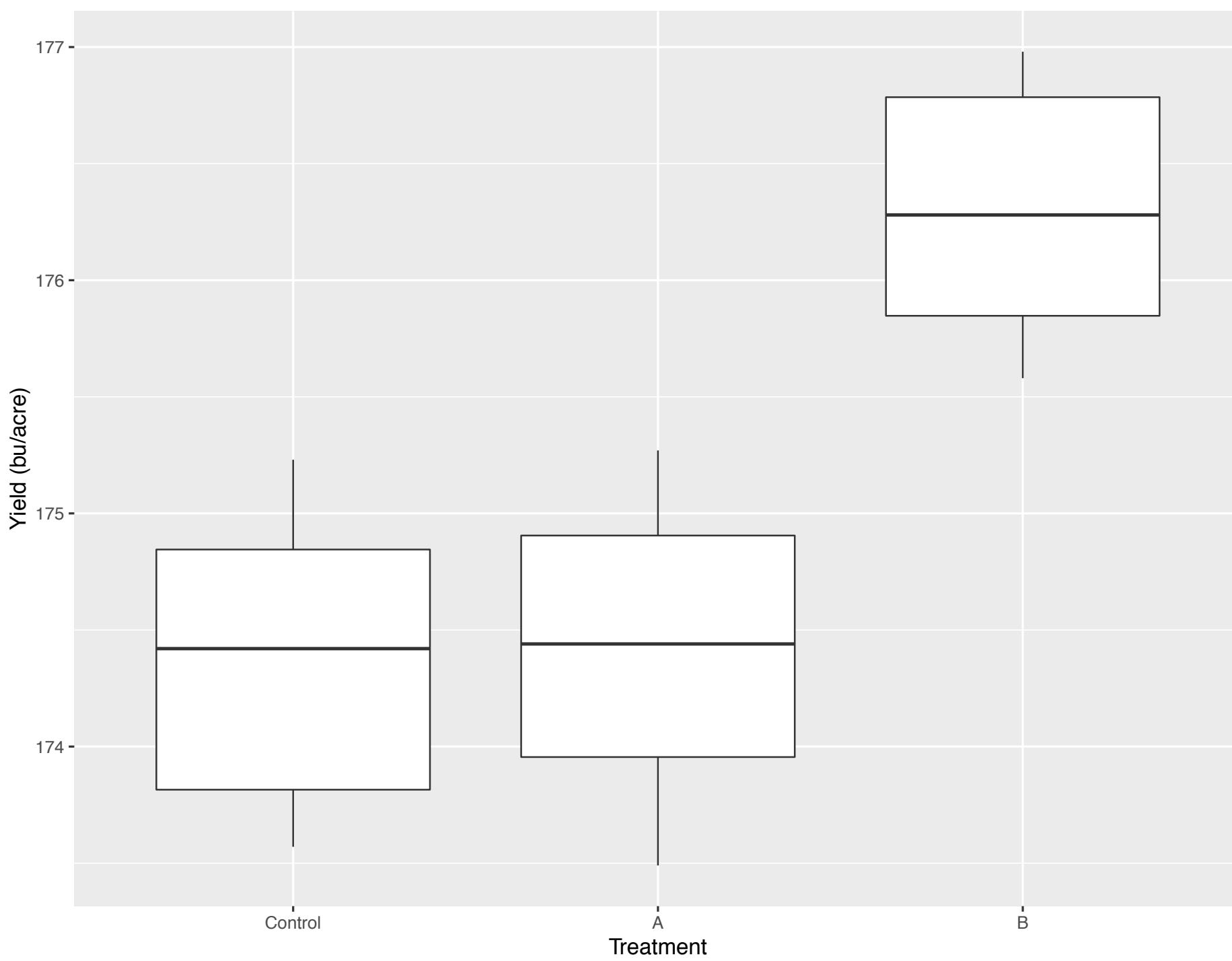


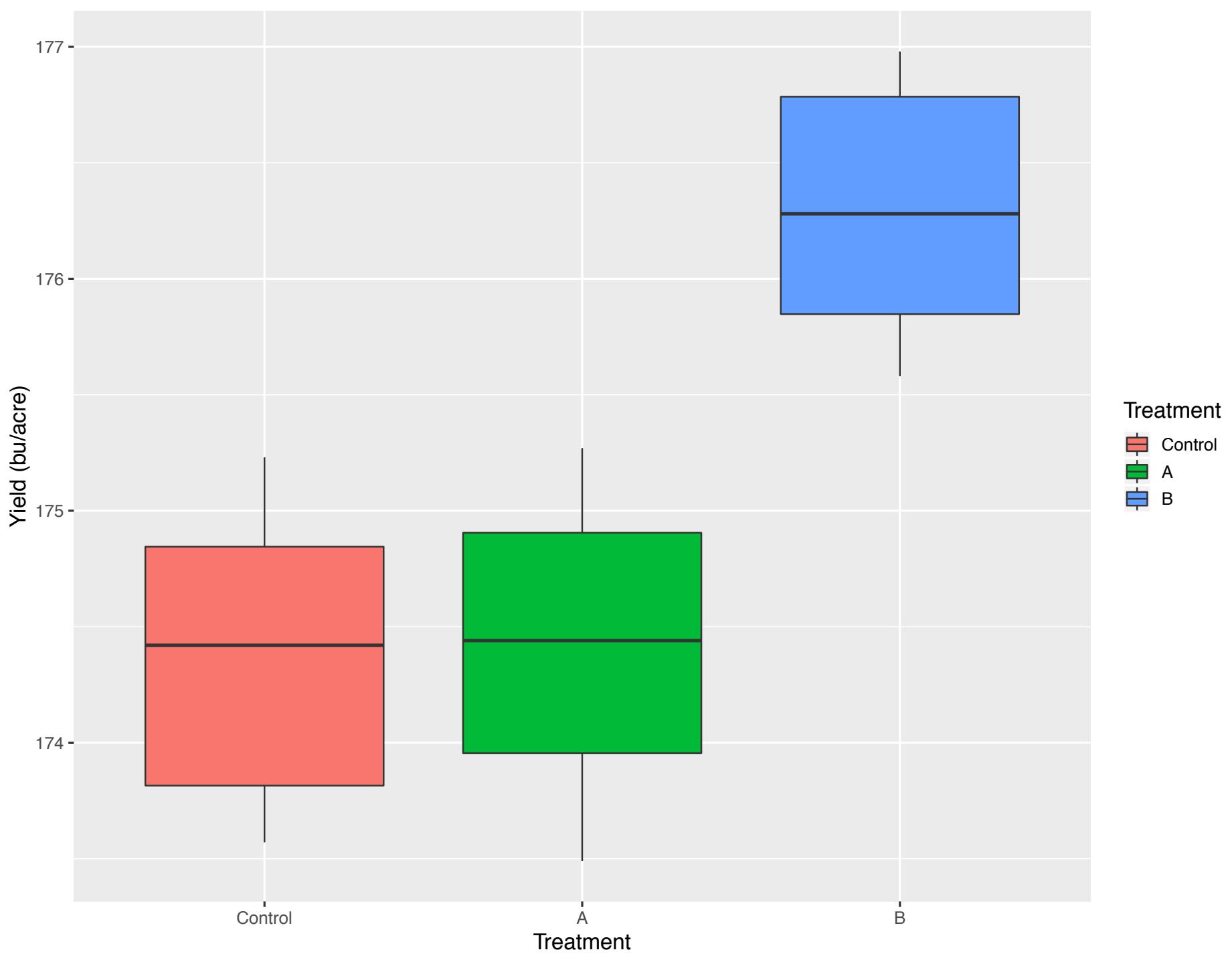


# How does ggplot2 work?

Treatment	Yield_bu_per_acre
Control	173.82
Control	174.23
Control	173.57
Control	173.61
Control	174.19
Control	173.80
Fungicide_A	173.98
Fungicide_A	174.27
Fungicide_A	173.61
Fungicide_A	173.88
Fungicide_A	174.17
Fungicide_A	173.49
Fungicide_B	175.98
Fungicide_B	175.58
Fungicide_B	175.75
Fungicide_B	175.88
Fungicide_B	175.68
Fungicide_B	175.95









# Let's Get Coding!

# Summary of Part 2b.

- `ggplot()` - initializes plot
- `aes()` - refers to your data
- `geom_()`- adds the specified geometry from data
- `theme()`- adjusts non-data components of plot
  - `element_()` – specifies display of non-data components
- Other packages can be used with `ggplot2`
  - `RColorBrewer`- can choose palette of your choice
  - `ggpubr`- can add p-values and significance levels
  - `plotrix`- can add standard error bars



# LEARN PRACTICE IMPROVE

BIGSTOCK

BIG  
STOCK

BIG  
STOCK

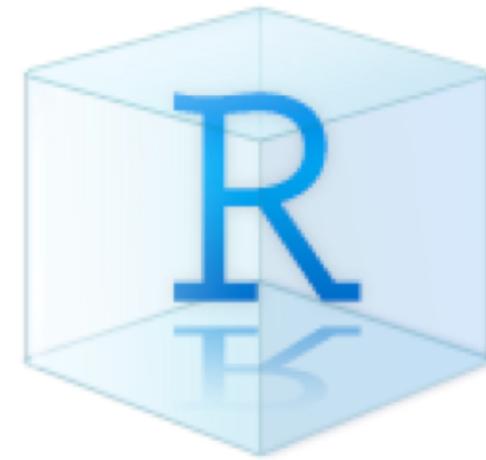


# Part 2c. Starting a Project

Katie Gold

## Part 2c: Start your own project

- Create a new R Project
- Create an R script inside a “scripts” folder
- Read in the dataset
- Move data into the “data” folder
- Save script and not workspace
- Re-start the R project



File Edit Code View Plots Session Build Debug Profile Tools Help

- New File
- New Project...**
- Open File... Ctrl+O
- Reopen with Encoding...
- Recent Files
- Open Project...
- Open Project in New Session...
- Recent Projects
- Import Dataset
- Save Ctrl+S
- Save As...
- Save with Encoding...
- Save All Ctrl+Alt+S
- Knit Document Ctrl+Shift+K
- Compile Report...
- Print...
- Close Ctrl+W
- Close All Ctrl+Shift+W
- Close All Except Current Ctrl+Alt+Shift+W
- Close Project
- Quit Session... Ctrl+Q

370:1 (Untitled) R Script

Console ~/IntroR\_2019/

```
R version 3.5.2 (2018-12-20) -- "Eggshell Igloo"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
```

```
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
```

```
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
```

```
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
```

&gt;

that we followed for 'yield.plot' to transform these

```
an_sev - se_sev,
an_sev + se_sev),
```

```
errors
```

```
tment),
```

```
point +
an_sev - se_sev,
an_sev + se_sev,
treatment),
```

Run Source

Environment History Connections Build Git

Global Environment

Environment is empty

Files Plots Packages Help Viewer

New Folder Delete Rename More

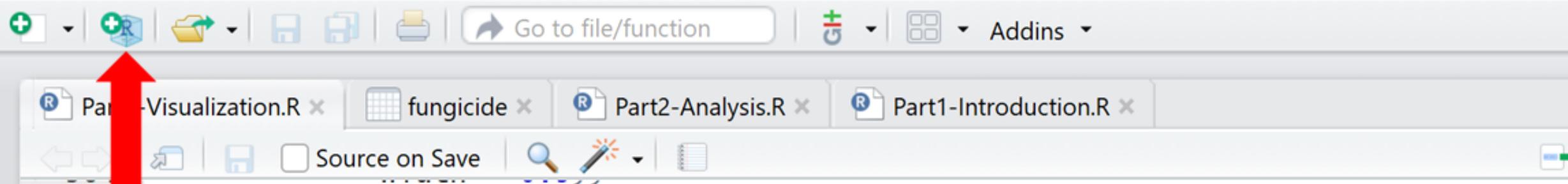
Home > IntroR\_2019

	Name	Size	Modified
<input type="checkbox"/>	..		
<input type="checkbox"/>	.gitignore	44 B	Feb 19, 2019, 3:19 PM
<input type="checkbox"/>	.Rhistory	13 KB	Mar 9, 2019, 2:54 PM
<input type="checkbox"/>	data		
<input type="checkbox"/>	docs		
<input type="checkbox"/>	IntroR_2019.Rproj	241 B	Mar 9, 2019, 2:54 PM
<input type="checkbox"/>	LICENSE	20.1 KB	Feb 19, 2019, 3:19 PM
<input type="checkbox"/>	Makefile	893 B	Mar 8, 2019, 2:53 PM
<input type="checkbox"/>	Part1-Introduction.R	19.6 KB	Mar 8, 2019, 2:53 PM
<input type="checkbox"/>	Part2-Analysis.R	13.2 KB	Mar 8, 2019, 2:53 PM
<input type="checkbox"/>	Part3-Visualization.R	12.8 KB	Mar 8, 2019, 4:01 PM



~/IntroR\_2019 - master - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help



```
365  
366 (m_se_plot_bar <- m_se_plot_bar +  
367     geom_errorbar(aes(ymin = mean_sev - se_sev,  
368                     ymax = mean_sev + se_sev),  
369                     width = 0.2))  
370  
371 # ### Point plot with standard errors  
372  
373 (m_se_plot_point <- m_se_plot +  
374     geom_point(aes(color = Treatment),  
375                  size = 3))  
376  
377 (m_se_plot_point <- m_se_plot_point +  
378     geom_errorbar(aes(ymin = mean_sev - se_sev,  
379                     ymax = mean_sev + se_sev,  
380                     color = Treatment)),
```

```
ar +  
mean_sev - se_sev,  
mean_sev + se_sev),  
))
```

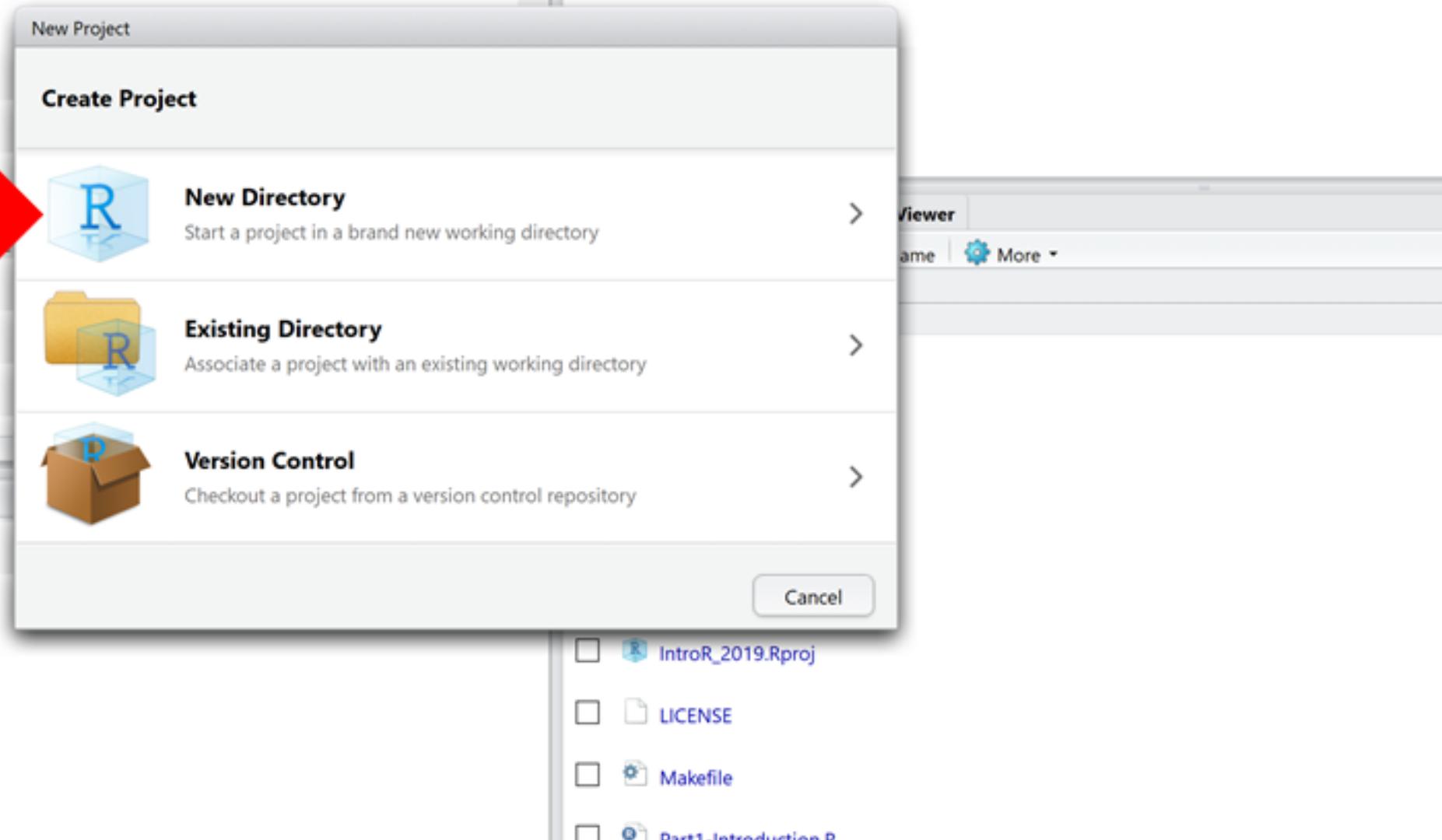
Environment is empty

d errors

```
+  
eatement),
```

```
_point +  
mean_sev - se_sev,  
mean_sev + se_sev,  
Treatment),  
))
```

ps that we followed for yield.plot.



rs

t),

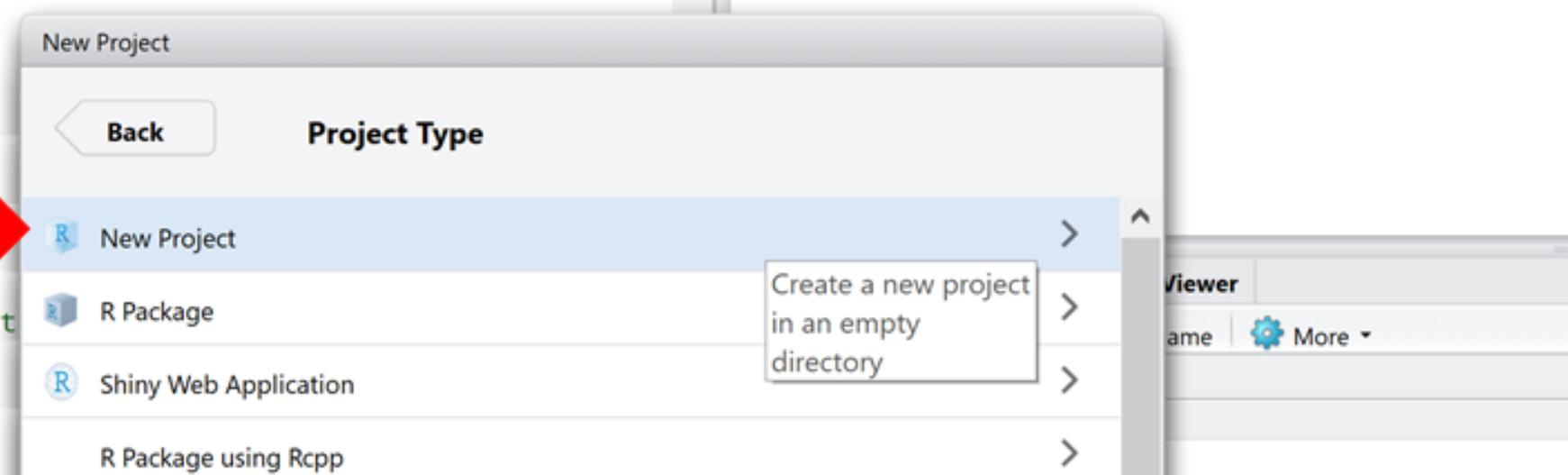
```
+  
ev - se_sev,  
ev + se_sev,  
ment),
```

t we followed for `yield.plot`

Igloo"  
atistical Computing

LY NO WARRANTY.  
ertain conditions.  
button details.

tributors.  
and

 IntroR\_2019.Rproj LICENSE Makefile

```
ev - se_sev,  
ev + se_sev),
```

Environment is empty

rs

t),

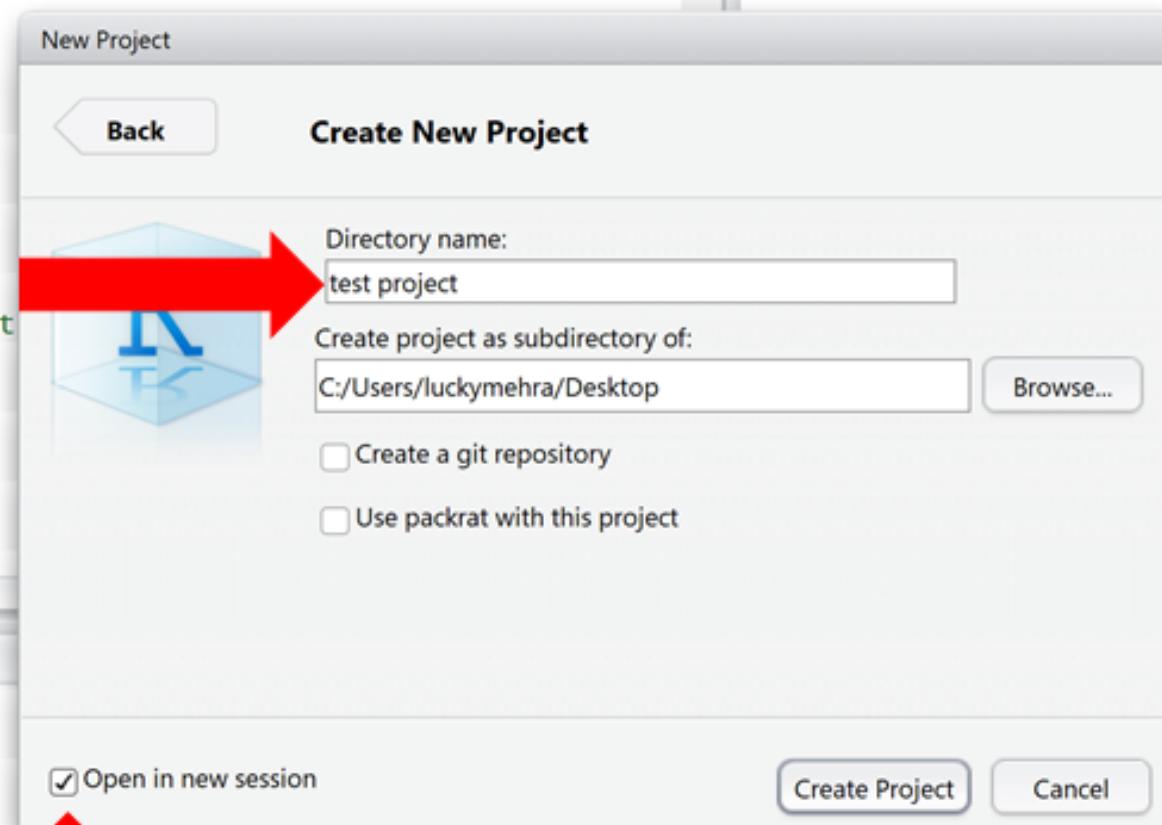
```
+  
ev - se_sev,  
ev + se_sev,  
ment),
```

t we followed for `yield.plot`

Igloo"  
atistical Computing

LY NO WARRANTY.  
ertain conditions.  
button details.

tributors.  
and  
s in publications.



-  IntroR\_2019.Rproj
-  LICENSE
-  Makefile
-  Part1-Introduction.R

C:/Users/luckymehra/Desktop/test project - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

+ - \_ Go to file/function Addins

R Script Ctrl+Shift+N

Create a new R

R Notebook

-12-20) -- "Eggshell Igloo"

The R Foundation for Statistical Computing  
64-mingw32/x64 (64-bit)

and comes with ABSOLUTELY NO WARRANTY.  
redistribute it under certain conditions.

'licence()' for distribution details.

ive project with many contributors.

'?' for more information and

'v' to cite R or R packages in publications.

some demos, 'help()' for on-line help, or  
an HTML browser interface to help.

Type 'q()' to quit R.

>

Environment History Connections

Import Dataset

Global Environment

Environment is empty

Files Plots Packages Help Viewer

New Folder Delete Rename More

C: > Users > luckymehra > Desktop > test project

Name	Size	Modified
test project.Rproj	218 B	Mar 9, 2019, 2:59 PM

## Packages

Resources

Twitter

For Windows UseRs



# References

## Packages

- [swirl](#) Interactive introduction to R so you can learn R, using R.
- [tidyverse](#) Core package to load packages dedicated to working with tidy data frames.

## Resources

- [R for Beginners](#) to build a strong foundation
- [RStudio Cheat Sheets](#) great visual maps for data analysis (highly recommend the dplyr, data import, and ggplot2 sheets!)
- [R style guide](#) to make sure your code is beautiful because when your code is easy to read, your future self will thank you.
- [UNL Statistics Help Desk](#) Statistical Cross-disciplinary Collaboration and Consulting Lab (SC3L)
- [CRAN task views](#) Grouping notable packages in specific topics such as Phylogenetics, Time Series, Experimental Design, etc.
- [Happy Git and GitHub for the useR](#) a long-form tutorial on working reproducibly in R and RStudio with Git and GitHub
- [Add P-values and Significance Levels to ggplots](#) an article on creating publication ready plots using ggpunr
- [Create colorful graphs in R](#) an article on creating colorful ggplots using RColorBrewer and plotly
- [Sampling, Data Management and Visualization](#) UNL Course led by Dr. Drew Tyre

## Twitter

Twitter is a great resource for getting help in R. You can browse the #rstats hashtag to see what's going on.

These accounts are great to follow to see what's going on in R

- [@RLangTip](#) tweets out useful tips and tricks to use R

# Please complete this brief feedback survey..

Intro to R    Home    Before the Workshop    Workshop ▾    Quiz    Links    About    Feedback Survey

Packages

Resources

Twitter

For Windows UseRs

## References

### Packages

- [swirl](#) Interactive introduction to R so you can learn R, using R.
- [tidyverse](#) Core package to load packages dedicated to working with tidy data frames.



# Contributors to this Workshop

- Zhian N. Kamvar
- Lucky Mehra
- O. William McClung
- Sydney Everhart
- Nikita Gambhir
- Katie Gold