Day 2: Data Visualization

github.com/mmckibben273/R-tist-Studio

Finish update

Product Solutions Resources Open Source Enterprise Pricing

Search or jump to... / Sign in Sign up

mmckibben273 / R-tist-Studio Public

Notifications Fork 0 Star 0

Code Issues Pull requests Actions Projects Security Insights

main 1 Branch 0 Tags Go to file Code

mmckibben273 Add files via upload 3acb5d7 · 4 hours ago 18 Commits

Cheatsheets	Delete Cheatsheets/filler	6 hours ago
Codes	Delete Codes/filler	6 hours ago
Datasets	Delete Datasets/filler	5 hours ago
Papers	Add files via upload	4 hours ago
Powerpoints	Create filler	6 hours ago
Poll Everywhere Login.pdf	Add files via upload	5 hours ago
README.md	Initial commit	6 hours ago

README

R-tist-Studio

About

Introductory class for highschool students to learn coding in RStudio

Readme
Activity
0 stars
1 watching
0 forks
Report repository

Releases

No releases published

Packages

No packages published

Languages

R 100.0%

R-tist-Studio / Codes /

Add file

mmckibben273 Add files via upload 97935e3 · 26 minutes ago History

Name	Last commit message	Last commit date
..		
Day1.Rmd	Update Day1.Rmd	3 days ago
Day2.Rmd	Add files via upload	26 minutes ago

github.com/mmckibben273/R-tist-Studio

Finish update

Product Solutions Resources Open Source Enterprise Pricing

Search or jump to... / Sign in Sign up

mmckibben273 / R-tist-Studio Public

Notifications Fork 0 Star 0

Code Issues Pull requests Actions Projects Security Insights

main 1 Branch 0 Tags Go to file Code

mmckibben273 Add files via upload 3acb5d7 · 4 hours ago 18 Commits

Cheatsheets	Delete Cheatsheets/filler	6 hours ago
Codes	Delete Codes/filler	6 hours ago
Datasets	Delete Datasets/filler	5 hours ago
Papers	Add files via upload	4 hours ago
Powerpoints	Create filler	6 hours ago
Poll Everywhere Login.pdf	Add files via upload	5 hours ago
README.md	Initial commit	6 hours ago

README

R-tist-Studio

About

Introductory class for highschool students to learn coding in RStudio

Readme
Activity
0 stars
1 watching
0 forks
Report repository

Releases

No releases published

Packages

No packages published

Languages

R 100.0%

R-tist-Studio / Datasets /

Add file

mmckibben273 Add files via upload 277c2c4 · 10 hours ago History

Name	Last commit message	Last commit date
..		
Gable et al. code:data	Add files via upload	3 days ago
Buckeye.csv	Add files via upload	10 hours ago
kong_mutation_data.txt	Add files via upload	3 days ago
tempandhr.txt	Add files via upload	3 days ago

R-tist-Studio / Datasets / Buckeye.csv

mmckibben273 Add files via upload 277c2c4 · 10 hours ago History

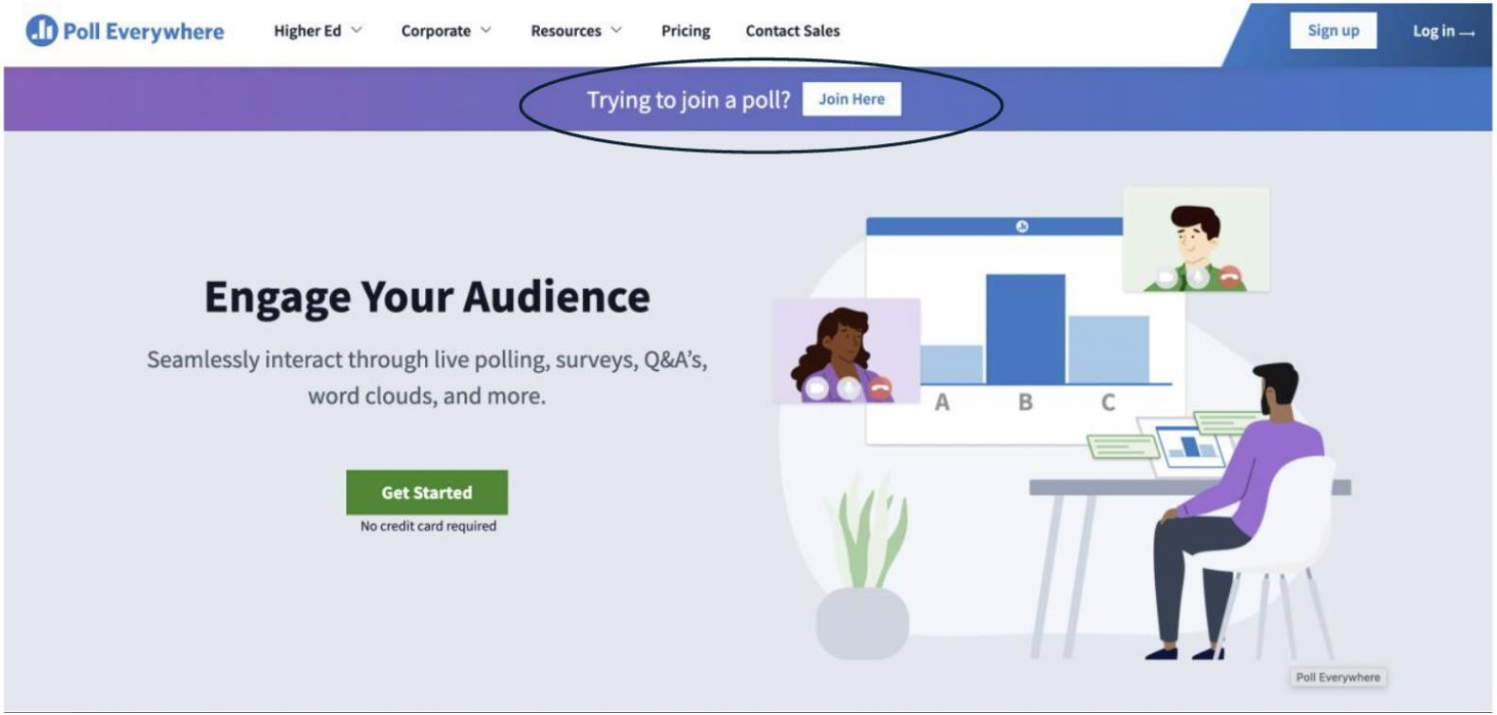
Code Blame 3.16 MB Code 55% faster with GitHub Copilot Raw

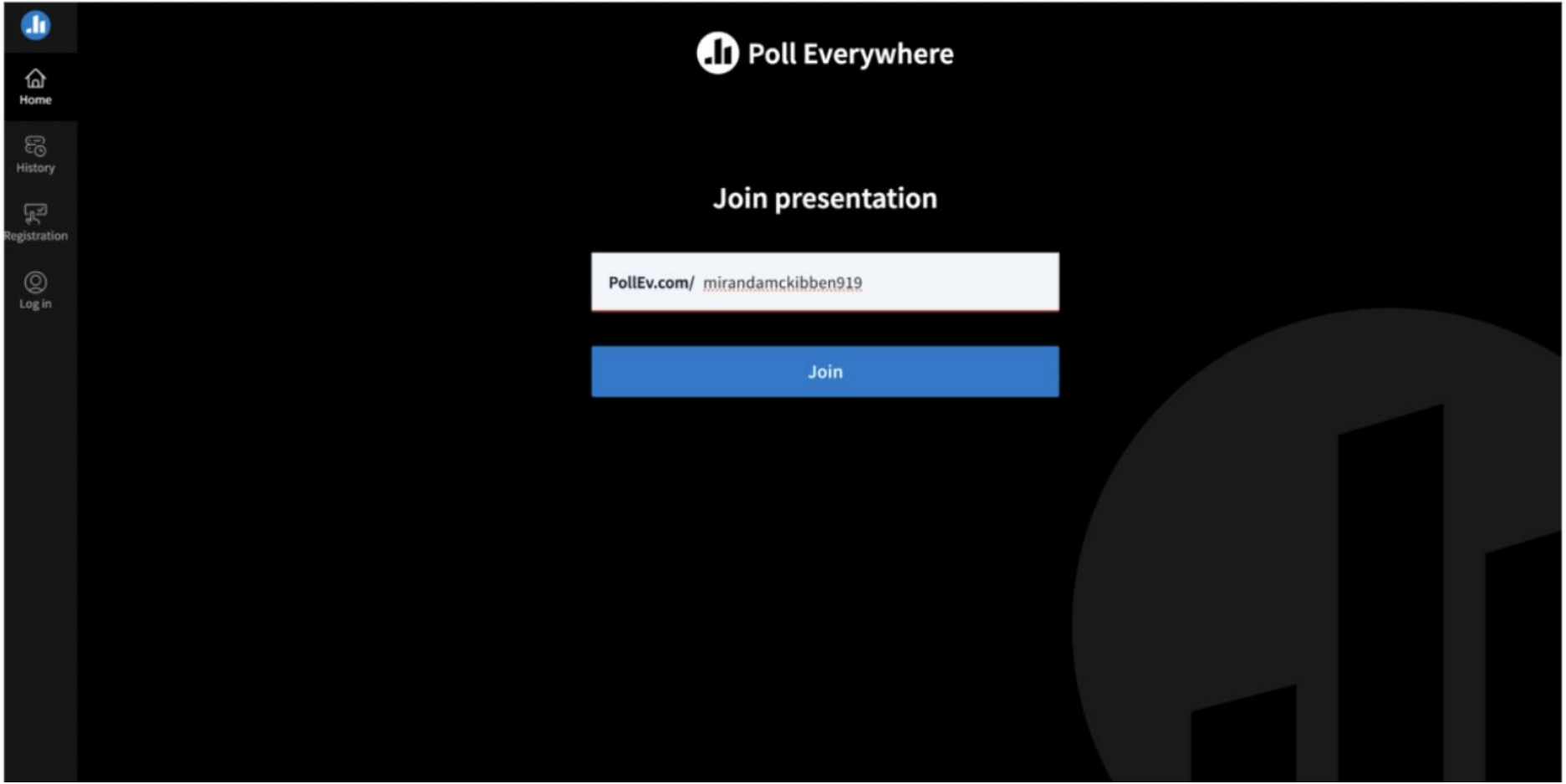
View raw

(Sorry about that, but we can't show files that are this big right now.)

Joining polleverywhere

- Go to www.polleverywhere.com




Home
History
Registration
Log in
Poll Everywhere
Join presentation
PollEv.com/ mirandamckibben919
Join

Home

History

Registration

Log in

Waiting for mirandamckibben919's presentation t

mirandamckibben919's presentation is underway. As soon as the activity is a

Poll Everywhere helps boost engagement during remote meetings, virtual tra

Welcome to mirandamckibben919's presentation!

Introduce yourself

Enter the screen name you would like to appear alongside your responses

Name
John Doe
8 / 50

Continue

Skip

Using a screen name allows the presenter and other participants to attach your screen name to your responses. You can change your screen name at any time.

How are you feeling today?

0

Goal of today's class

- Identify whether a variable is continuous or categorical, qualitative or quantitative
- Practice making plots and figures in RStudio

You are given an experimental dataset and told to create a figure for it. What is the first information you should figure out?

- Whether the data is quantitative or qualitative
- The independent and dependent variables
- What kind of figure you will make
- Whether data is continuous or categorical
- None of the above

You are given an experimental dataset and told to create a figure for it. What is the first information you should figure out?

Whether the data is quantitative or qualitative 0%

The independent and dependent variables 0%

What kind of figure you will make 0%

Whether data is continuous or categorical 0%

None of the above 0%

You are given an experimental dataset and told to create a figure for it. What is the first information you should figure out?

Whether the data is quantitative or qualitative 0%

The independent and dependent variables 0%

What kind of figure you will make 0%

Whether data is continuous or categorical 0%

None of the above 0%

The first step in graphing is identifying your variables

Independent Variable (x-axis)

A variable whose value does not depend on any variable (independent)

Think of it as the “cause” in experimental research

Example: abiotic factors (temperature, salinity, food, O2)

Dependent Variable (y-axis)

A variable whose value does depend on the manipulation of another variable (dependent)

Think of it as the “effect” in experimental research

Example: biotic responses (movement, growth, reproduction)

Step 2: Are your variables qualitative or quantitative?

Qualitative variables

- Descriptions, language or text rather than numbers
- Usually needs to be categorized
- Examples:
  - How do students feel about each course?
  - What do the tree leaves look like over time?

Quantitative variables

- Numerical
- Quantifiable, more precise
- Examples:
  - How many students are in each course?
  - How does tree height change over time?

Step 3: Is your independent variable categorical or continuous?

Categorical

Data can be assigned to discrete groups or categories

Organism based: species, color, leaf shape, etc.

Treatment based: fertilized vs. unfertilized, experimental drug vs. placebo, etc.

Continuous

Data is numeric with an infinite number of possible values

Organism based: height, weight, age, etc.

Treatment based: nutrient concentration, time, amount of daylight, temperature, etc.

You want to use a dataset that measures how much snow Columbus, Ohio had over the past 20 years. What is your independent variable?

Amount of snow 0%

Year 0%

Columbus 0%

None of the above 0%

You want to know how many football games the Ohio State Buckeyes win on average in a season. What kind of data are you collecting?

Qualitative

0%

Quantitative

0%

You want to know how long it typically takes for Donatos to deliver a pizza. Is your independent variable continuous or categorical?

Categorical

0%

Continuous

0%

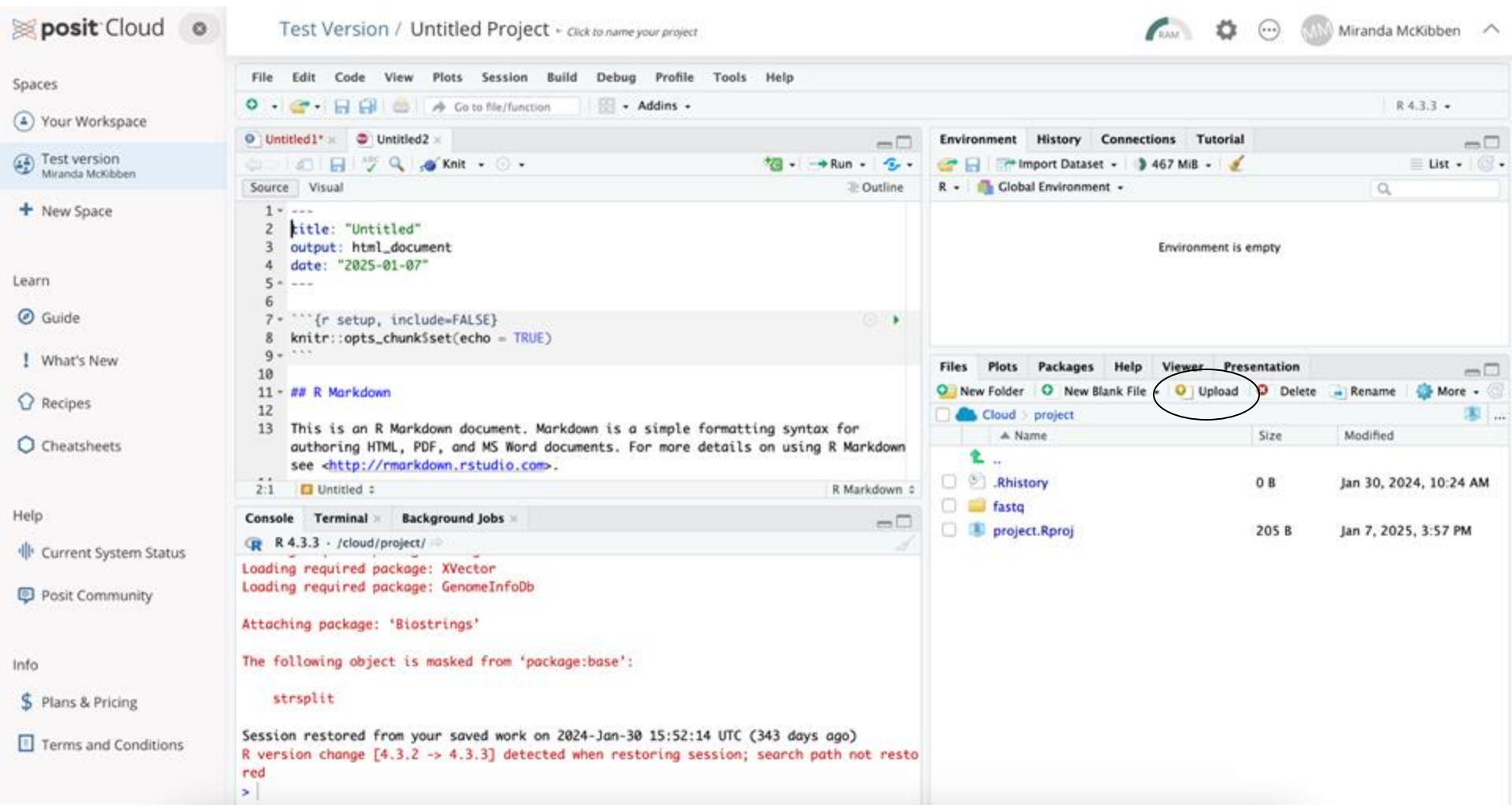
Types of plots to use for different types of data

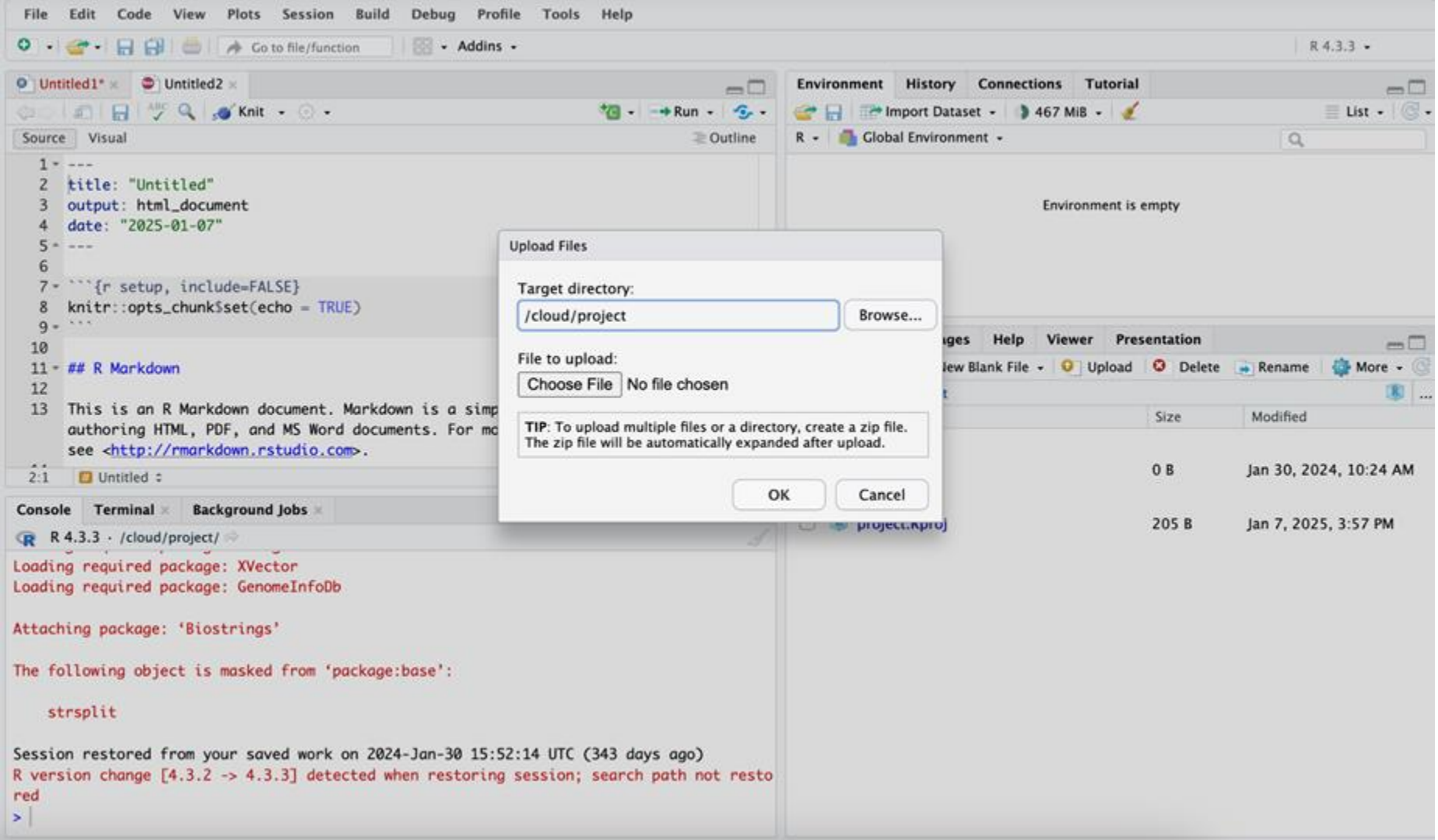
Categorical

Bar chart
Pie chart
Box and Whisker Plot

Continuous

Line graph
Scatter Plot
Histogram


posit Cloud
Test Version / Untitled Project
Click to name your project
Miranda McKibben
Spaces
Your Workspace
Test version
Miranda McKibben
New Space
Learn
Guide
What's New
Recipes
Cheatsheets
Help
Current System Status
Posit Community
Info
Plans & Pricing
Terms and Conditions
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
R 4.3.3
Untitled1*
Untitled2
Knit
Run
Source
Visual
Outline
1 ---
2 title: "Untitled"
3 output: html_document
4 date: "2025-01-07"
5 ---
6
7 ```{r setup, include=FALSE}
8 knitr::opts_chunk$set(echo = TRUE)
9 ```
10
11 ## R Markdown
12
13 This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.
2:1
Untitled
R Markdown
Console
Terminal
Background Jobs
R 4.3.3 · /cloud/project/
Loading required package: XVector
Loading required package: GenomeInfoDb
Attaching package: 'Biostrings'
The following object is masked from 'package:base':
strsplit
Session restored from your saved work on 2024-Jan-30 15:52:14 UTC (343 days ago)
R version change [4.3.2 -> 4.3.3] detected when restoring session; search path not restored
>
Environment
History
Connections
Tutorial
Import Dataset
467 MiB
List
R
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
Presentation
New Folder
New Blank File
Upload
Delete
Rename
More
Cloud
project
Name
Size
Modified
..
.Rhistory
0 B
Jan 30, 2024, 10:24 AM
fastq
project.Rproj
205 B
Jan 7, 2025, 3:57 PM


File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
R 4.3.3
Untitled1*
Untitled2
Knit
Run
Source
Visual
Outline
1 ---
2 title: "Untitled"
3 output: html_document
4 date: "2025-01-07"
5 ---
6
7 ```{r setup, include=FALSE}
8 knitr::opts_chunk$set(echo = TRUE)
9 ```
10
11 ## R Markdown
12
13 This is an R Markdown document. Markdown is a simp
authoring HTML, PDF, and MS Word documents. For mo
see <http://rmarkdown.rstudio.com>.
2:1
Untitled
Environment
History
Connections
Tutorial
Import Dataset
467 MiB
List
R
Global Environment
Environment is empty
Upload Files
Target directory:
/cloud/project
Browse...
File to upload:
Choose File
No file chosen
TIP: To upload multiple files or a directory, create a zip file.
The zip file will be automatically expanded after upload.
OK
Cancel
Help
Viewer
Presentation
New Blank File
Upload
Delete
Rename
More
Size
Modified
0 B
Jan 30, 2024, 10:24 AM
205 B
Jan 7, 2025, 3:57 PM
Console
Terminal
Background Jobs
R 4.3.3 · /cloud/project/
Loading required package: XVector
Loading required package: GenomeInfoDb
Attaching package: 'Biostrings'
The following object is masked from 'package:base':
strsplit
Session restored from your saved work on 2024-Jan-30 15:52:14 UTC (343 days ago)
R version change [4.3.2 -> 4.3.3] detected when restoring session; search path not restored
>

```{r }
rm(list=ls())
data <- airquality #This line calls the airquality set and gives it the name "data"
head(data) #This shows a preview of the dataset
```

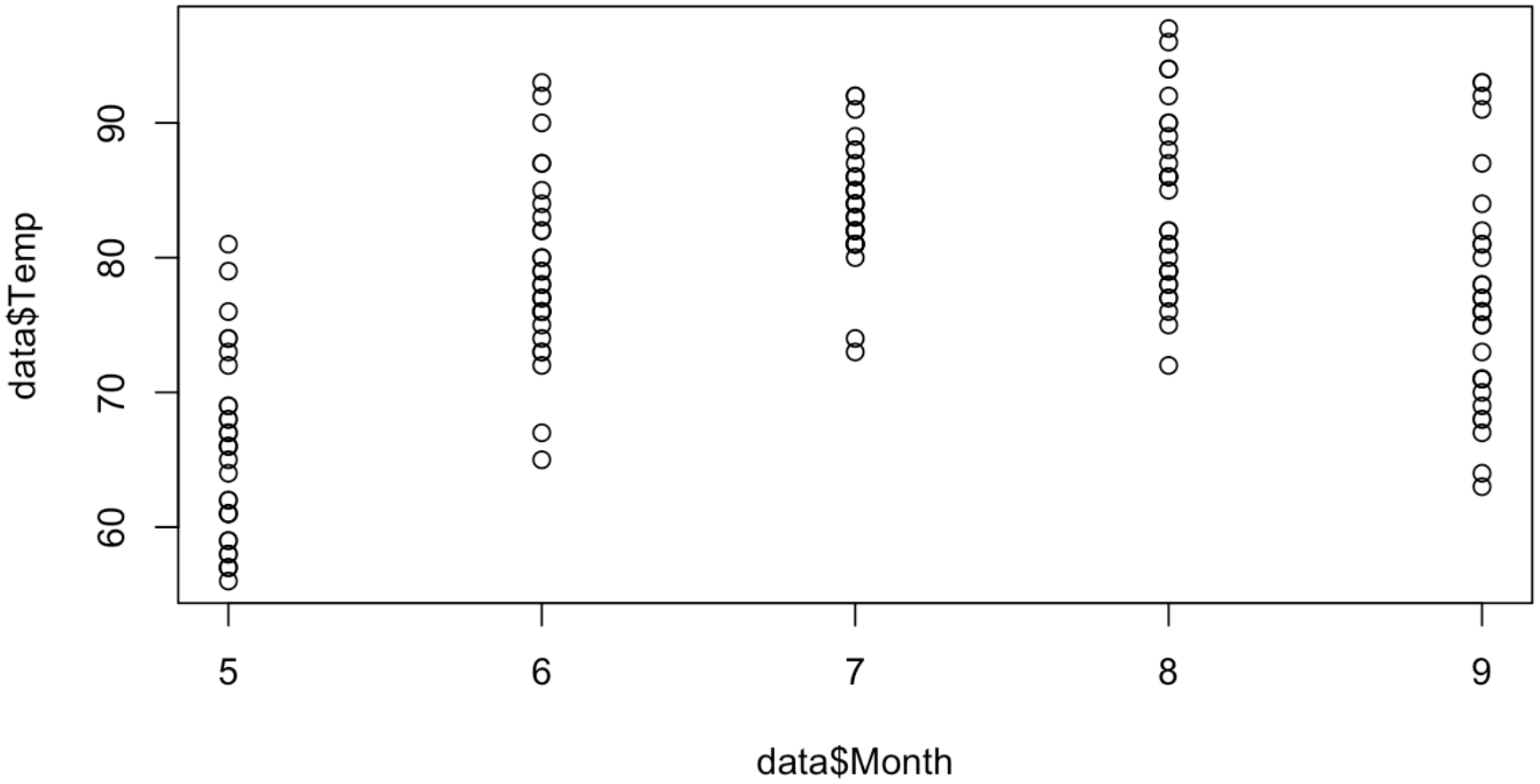
Description: df [6 × 6]

	Ozone <int>	Solar.R <int>	Wind <dbl>	Temp <int>	Month <int>	Day <int>
1	41	190	7.4	67	5	1
2	36	118	8.0	72	5	2
3	12	149	12.6	74	5	3
4	18	313	11.5	62	5	4
5	*NA*	*NA*	14.3	56	5	5
6	28	*NA*	14.9	66	5	6

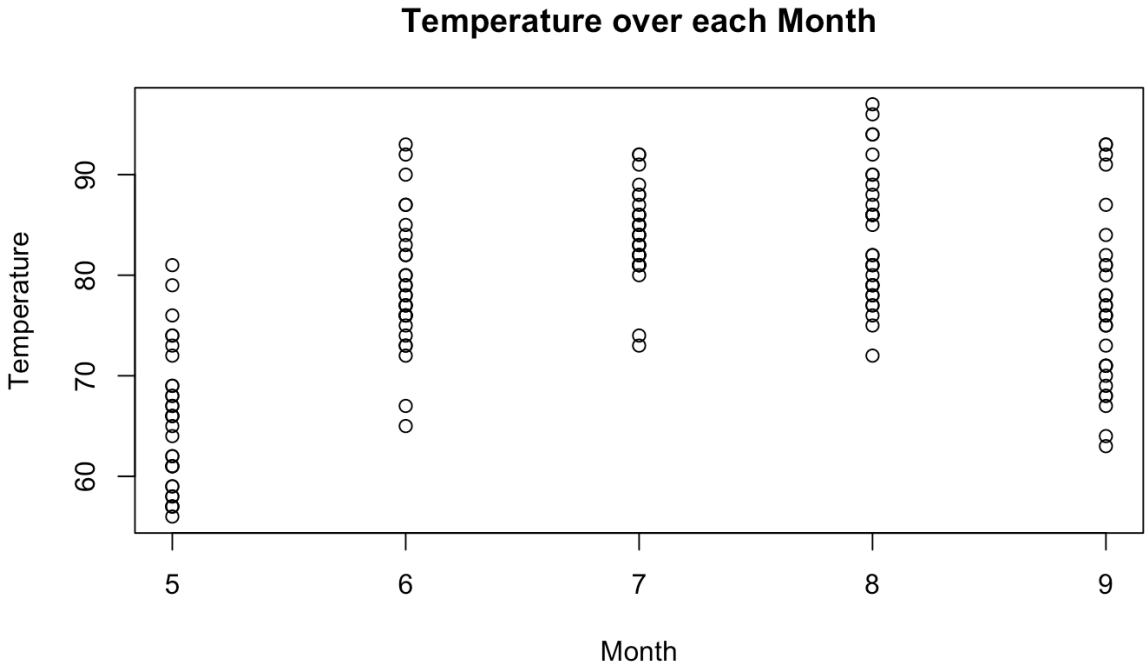
6 rows

Let's create a simple plot of the dataset. To plot data, we need to tell RStudio what we want the x and y axis to be.

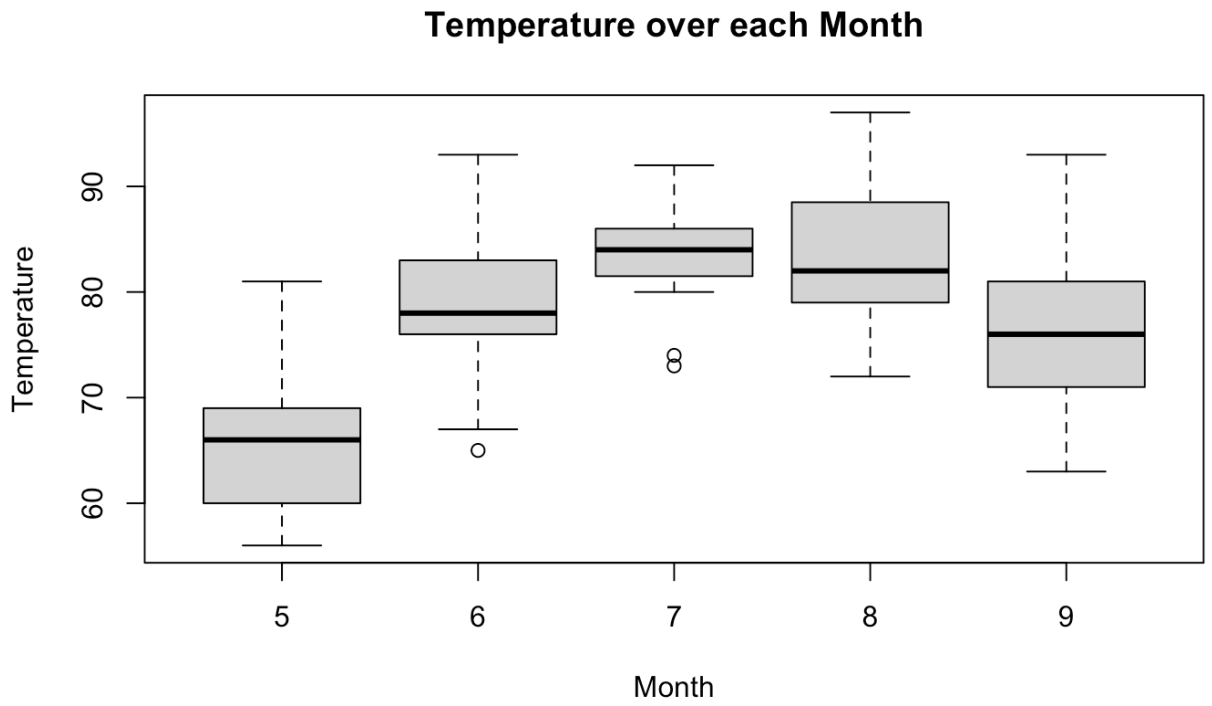
```{r}
plot(data$Month,data$Temp)
```


data$Temp
60
70
80
90
5
6
7
8
9
data$Month

```{r}
plot(data$Month,data$Temp, xlab = "Month", ylab = "Temperature",main = "Temperature over each Month")

boxplot(Temp ~ Month, data=data, xlab = "Month", ylab = "Temperature",main = "Temperature over each Month") #Notice the input for a boxplot is dependent~independent variable
```



```{r}
plot(data$Month,data$Temp, xlab = "Month", ylab = "Temperature",main = "Temperature over each Month")

boxplot(Temp ~ Month, data=data, xlab = "Month", ylab = "Temperature",main = "Temperature over each Month") #Notice the input for a boxplot is dependent~independent variable
```

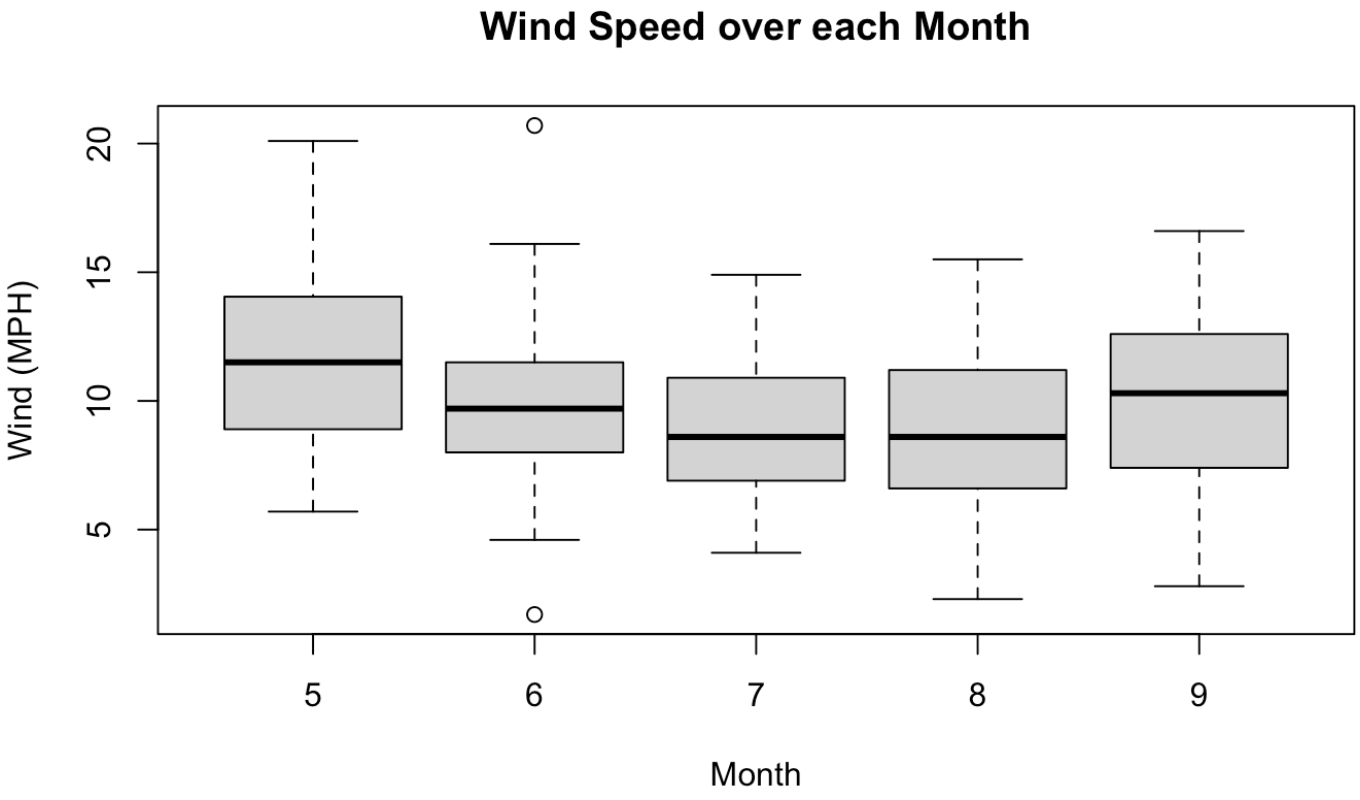


Now make your own boxplot for wind speed each month.

```{r}

```

Now make your own boxplot for wind speed each month.

```{r}
boxplot(Wind ~ Month, data=data, xlab = "Month", ylab = "Wind (MPH)",main = "Wind Speed over each Month") #Notice the input for a boxplot is dependent~independent variable
```



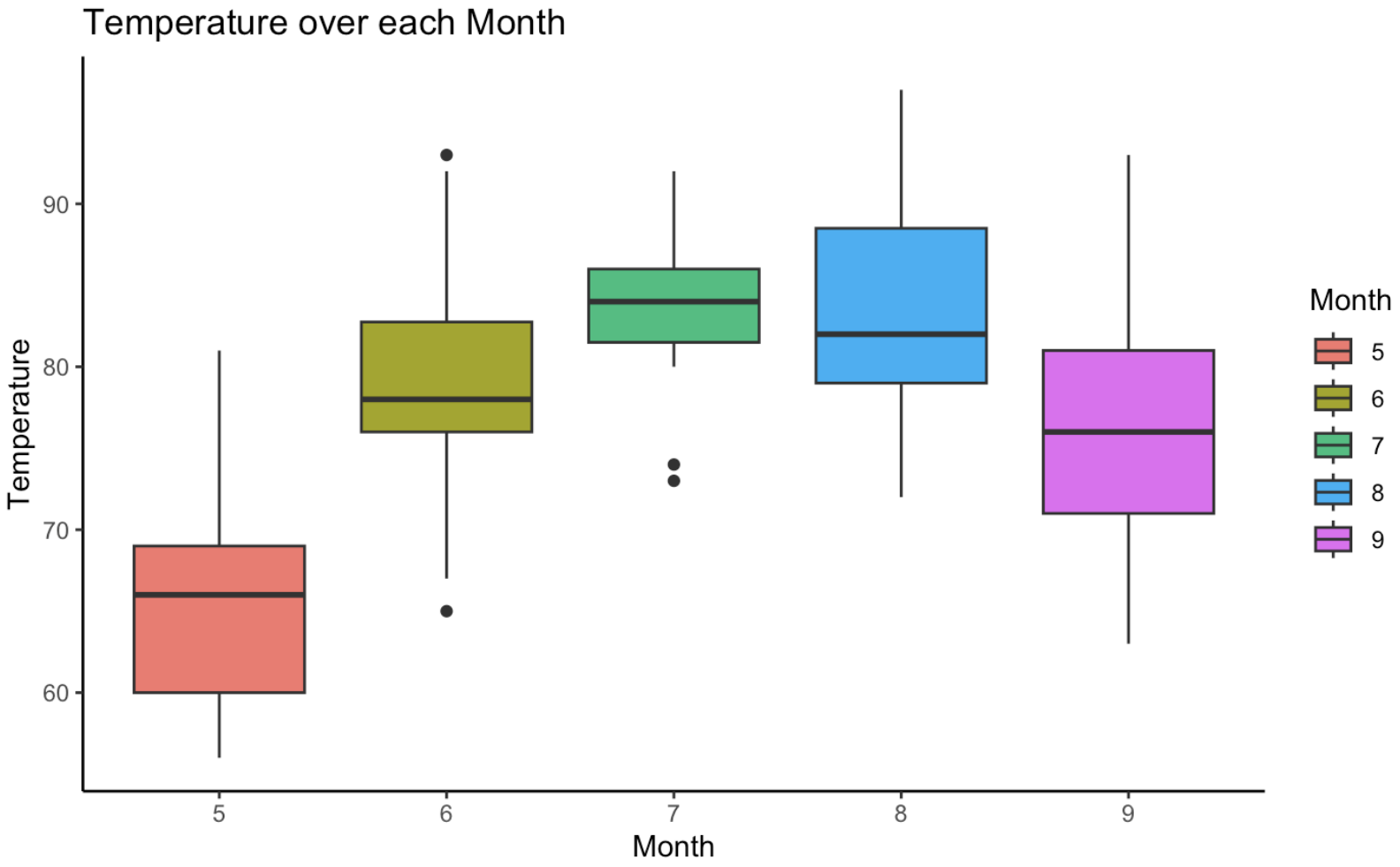
Have you made a box plot?

Yes

0%

No

0%

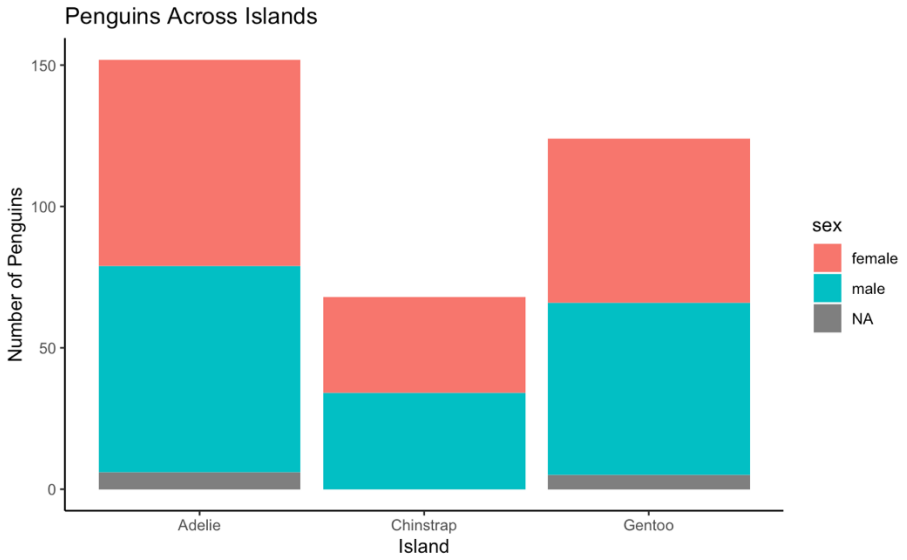

Temperature over each Month
Temperature
90
80
70
60
5
6
7
8
9
Month
Month
5
6
7
8
9

Imitating the code above, run your own code creating a bar graph for the number of penguins in each species sampled and how many of each sex there were.

```{r}

```

Imitating the code above, run your own code creating a bar graph for the number of penguins in each species sampled and how many of each sex there were.

```{r}
ggplot(data1) + #This line tells RStudio what dataset you are pulling data from
  aes(x = species, fill = sex)+ #The aesthetics command tells ggplot what variable from the dataset to plot on the x axis
  geom_bar()+ #What type of figure/plot to create, in this case a bar plot with geom_bar()
  ylab("Number of Penguins") + # y label
  xlab("Island") + # x label
  ggtitle("Penguins Across Islands")+ #Title of dataset
  theme_classic() #Theme changes the background from a grid to clear white. Feel free to explore how other themes (check Day 2 powerpoint for list of possible themes or look it up on your own)
```



Have you made a bar graph?

Yes

0%

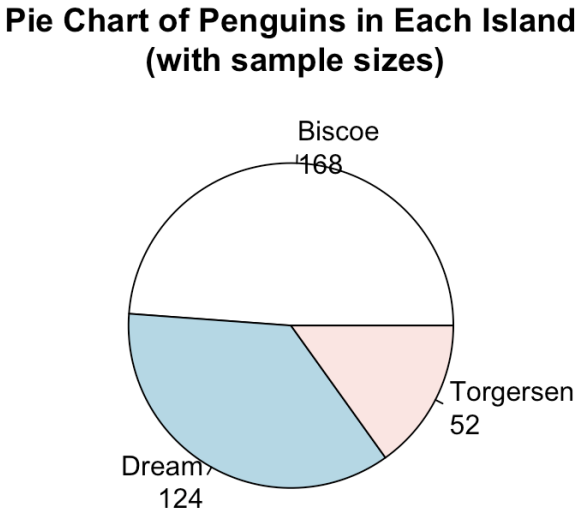
No

0%

Create a pie chart showing how many penguins were sampled on each island.

```{r}

```

```{r}
mytable <- table(data1$island) #This command pulls out the column we will be making a pie chart with and makes it into a table format
head(mytable)
lbls <- paste(names(mytable), "\n", mytable, sep="") # The paste() command tells RStudio to combine strings of characters and/or numbers. First, we pull the names of the columns from the dataset mytable. Then the command "/n" attaches the number in the column to the name. We then put the dataset we are pulling the labels from, mytable, and what separates columns, in this case spaces, which are represented by sep = "".
pie(mytable, labels = lbls,
   main="Pie Chart of Penguins in Each Island\n (with sample sizes)")
```



Have you made a pie chart?

Yes

0%

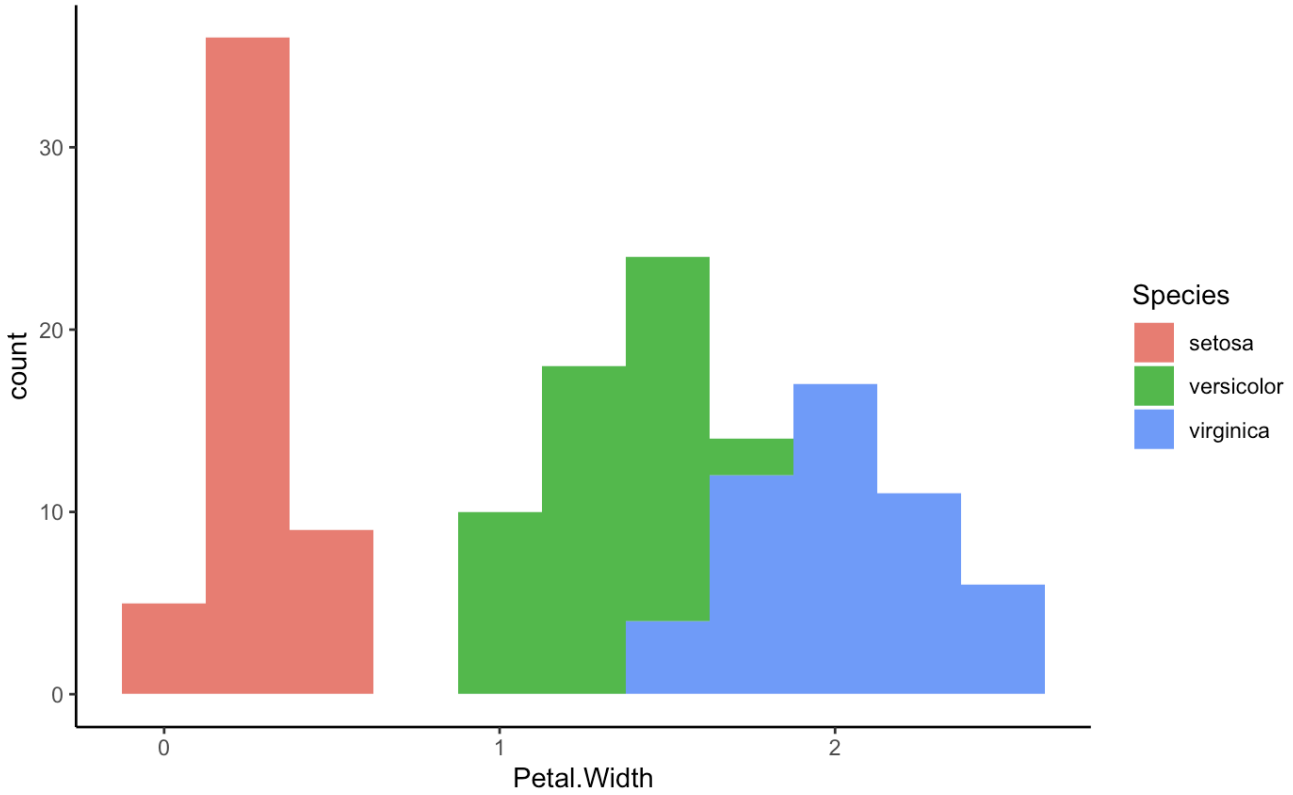
No

0%

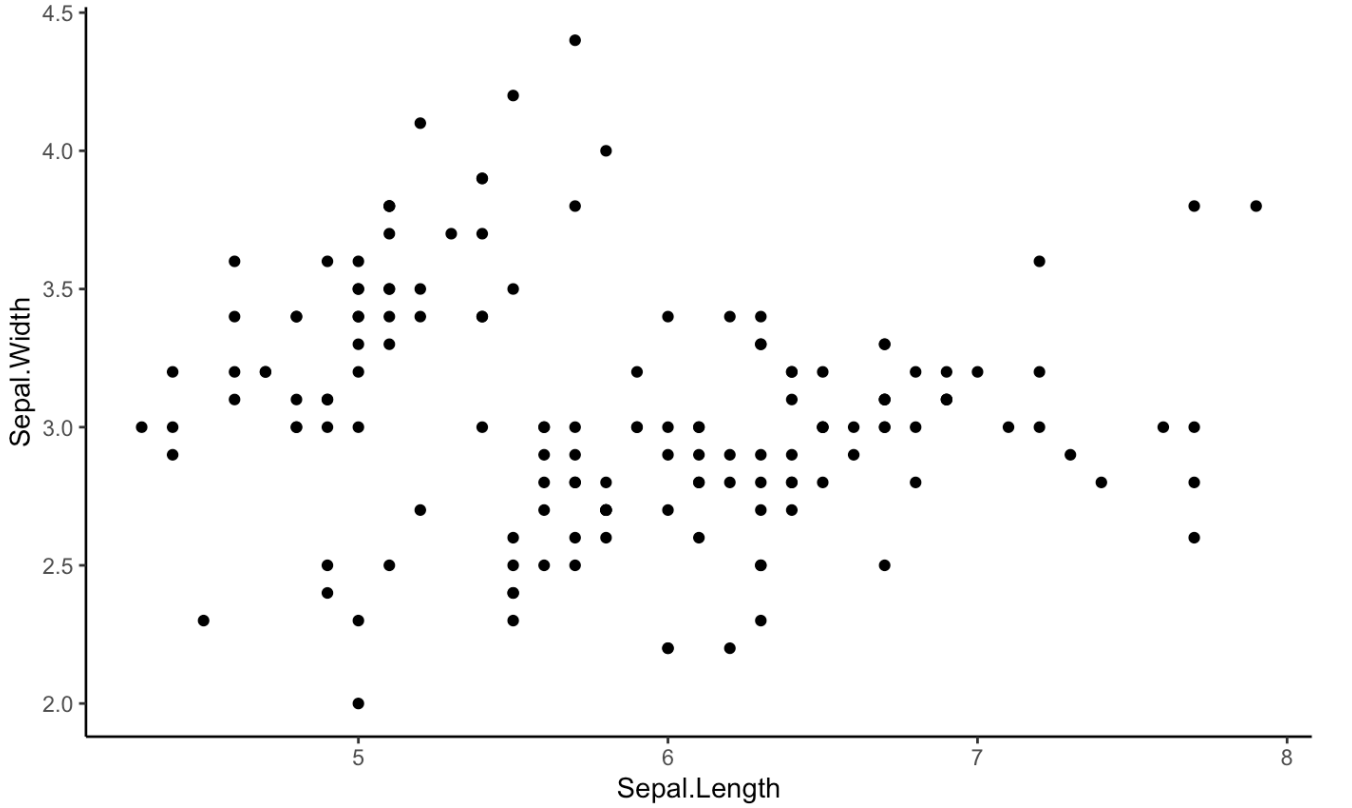
Create your own histogram looking at petal width across species.

```{r}

```

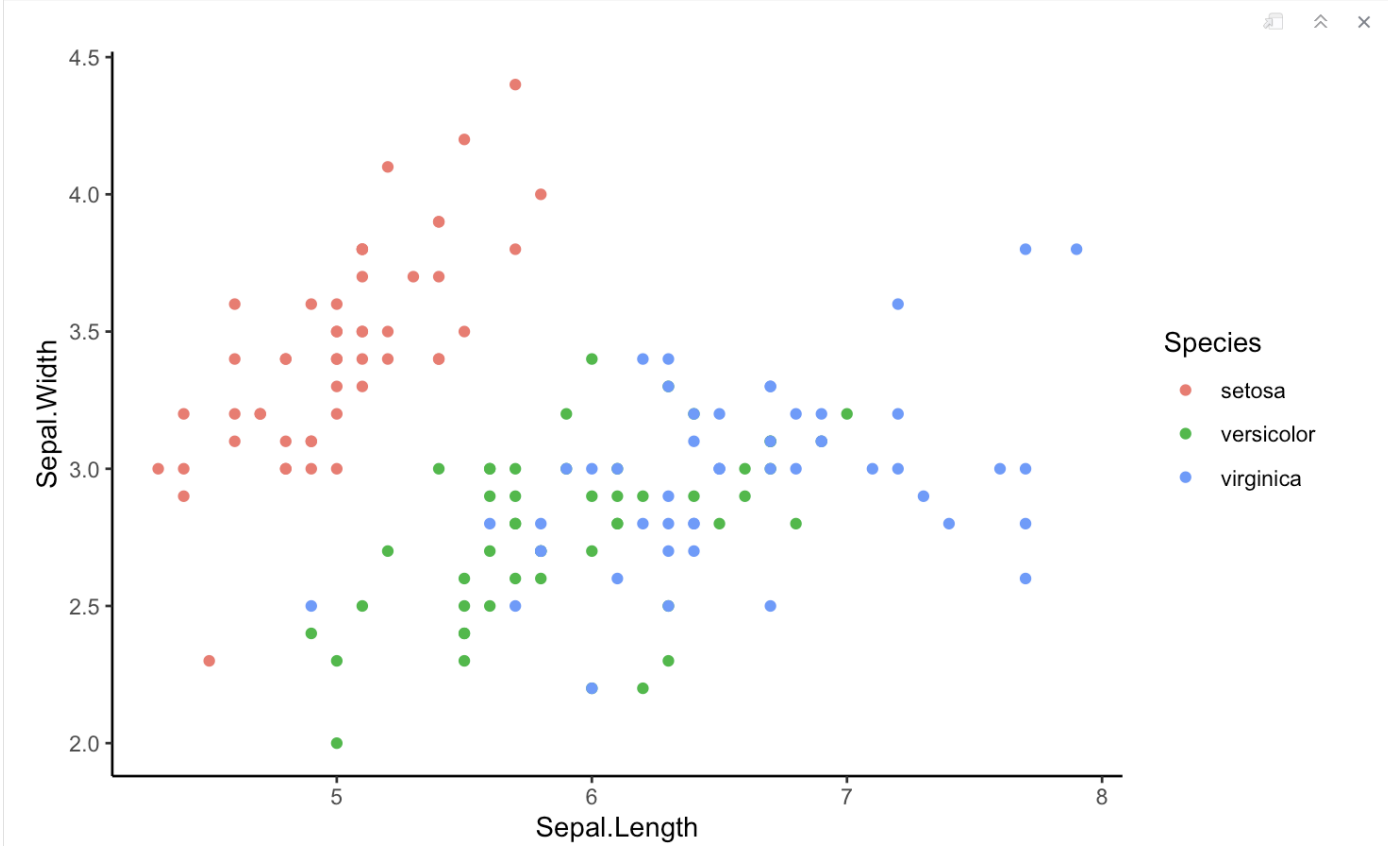
```{r}
ggplot(data2, aes(x = Petal.Width, fill=Species)) +
  geom_histogram(binwidth = 0.25)+ #Try changing the bin width and rerunning the code to see how it changes your visualization. When deciding binwidths for your histograms, make sure to think about the range/length of your x axis.
  theme_classic()
```

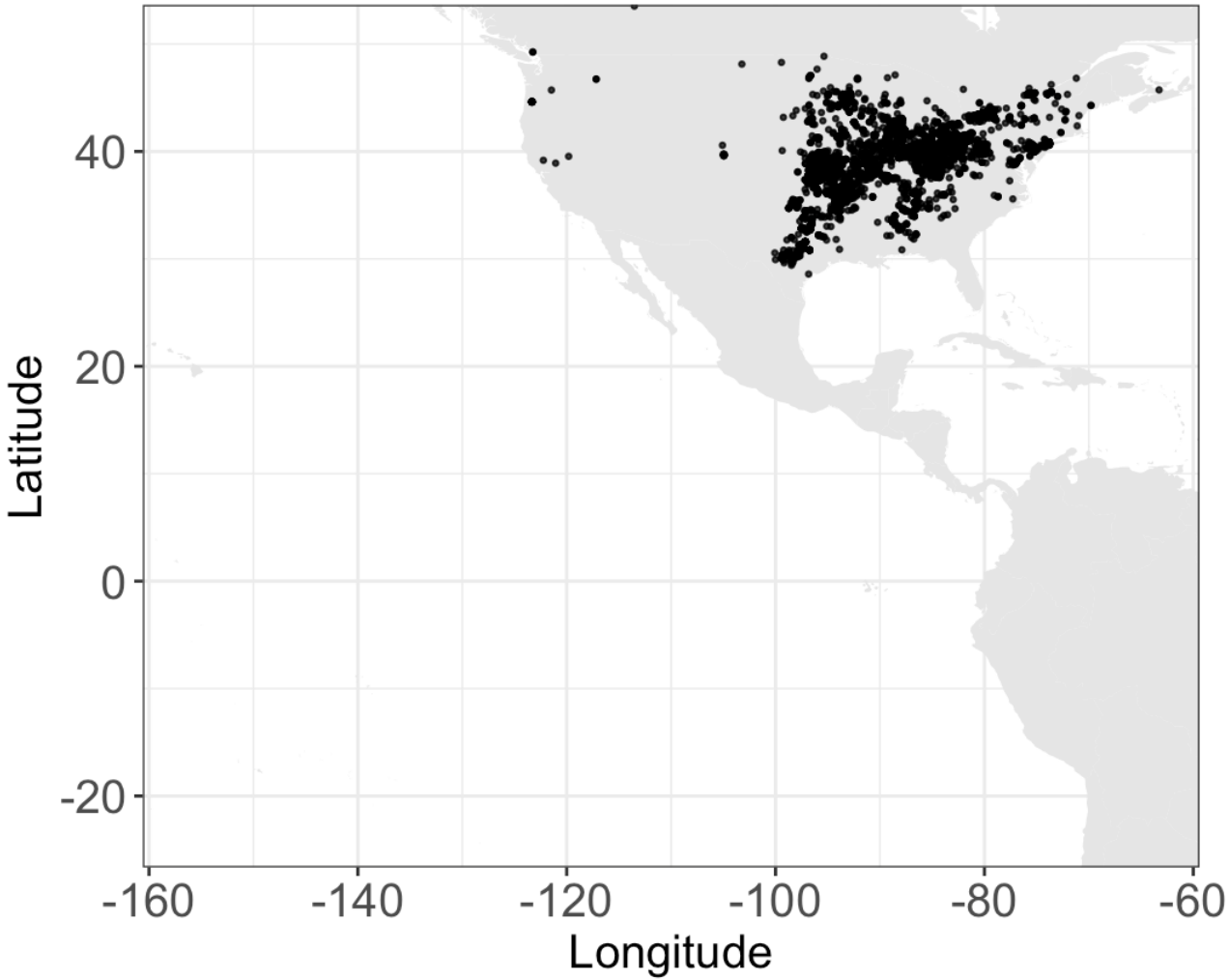


```{r}
ggplot(data = data2) +
  aes(x = Sepal.Length, y = Sepal.Width) +
  geom_point() +
  theme_classic()

```

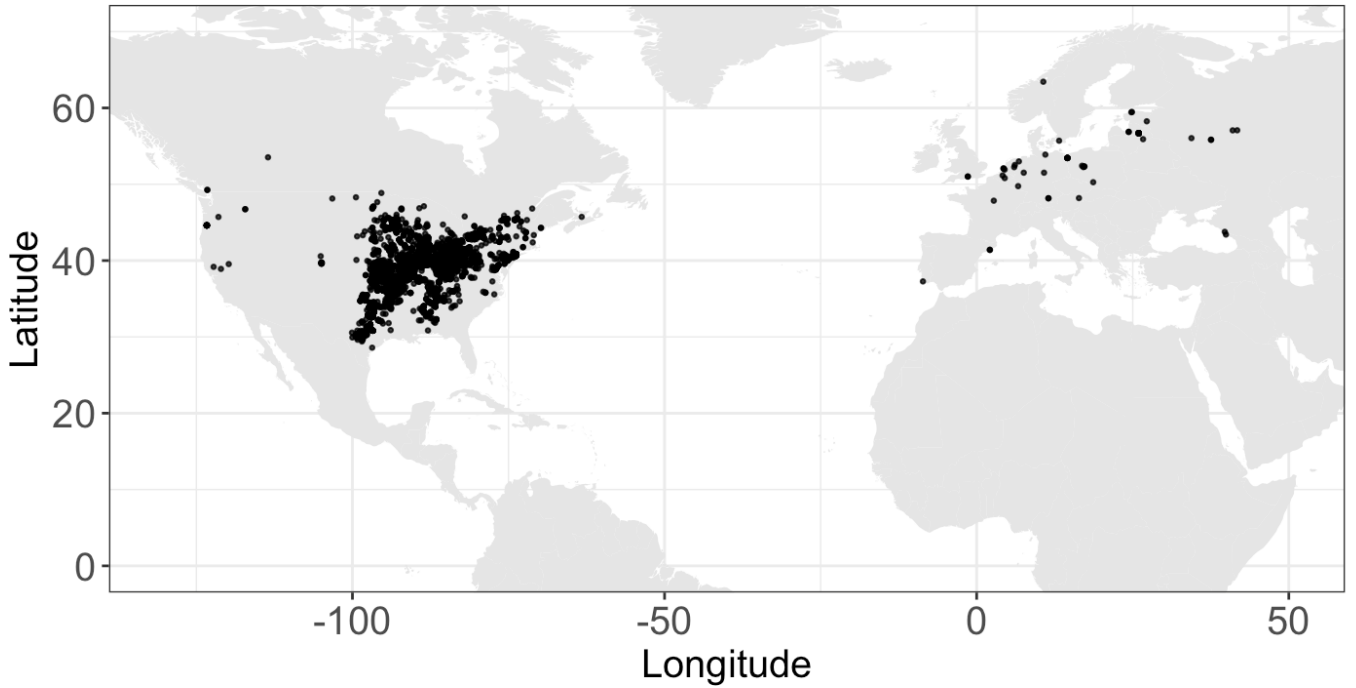


```{r}
ggplot(data = data2) +
  aes(x = Sepal.Length, y = Sepal.Width, color=Species) + #In this line we tell ggplot to color the points by the flower species
  geom_point() +
  theme_classic()
```




40
20
0
-20
Latitude
-160
-140
-120
-100
-80
-60
Longitude

```
  coord_quickmap(xlim=c(-130,50),ylim=c(0,70)) + #Play with the latitude and longitude to make the graph more centered
  theme_bw()+theme(legend.position="right",legend.title=element_blank())+
  theme(axis.text=element_text(size=15))+
  theme(axis.title=element_text(size=15))
plot(gg1)
```



How could I improve this class for you? Is there anything you would like us to review more in depth next week?

Nobody has responded yet.

Hang tight! Responses are coming in.