Day 2: Identifying types of variables and practicing basic plots in R

R-tist-Studio / Codes /

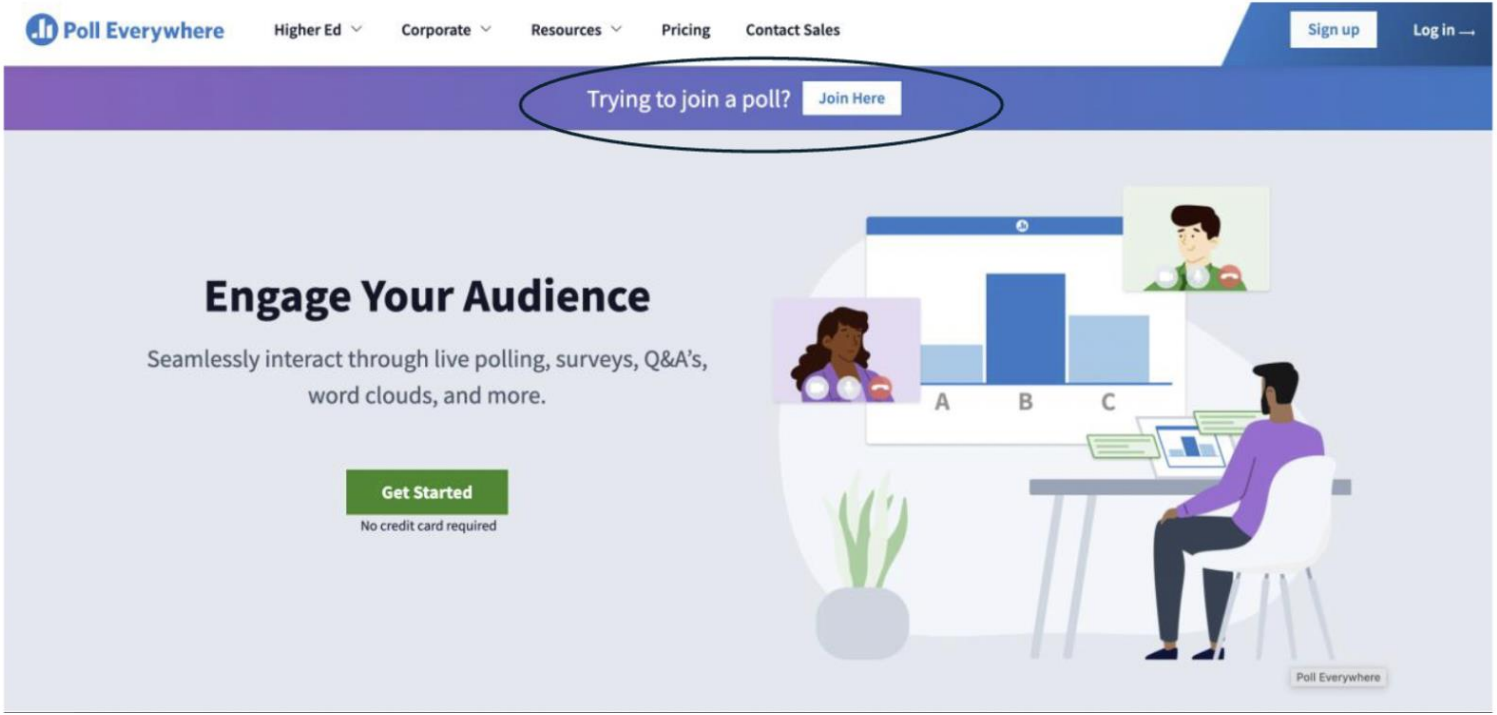
Add file

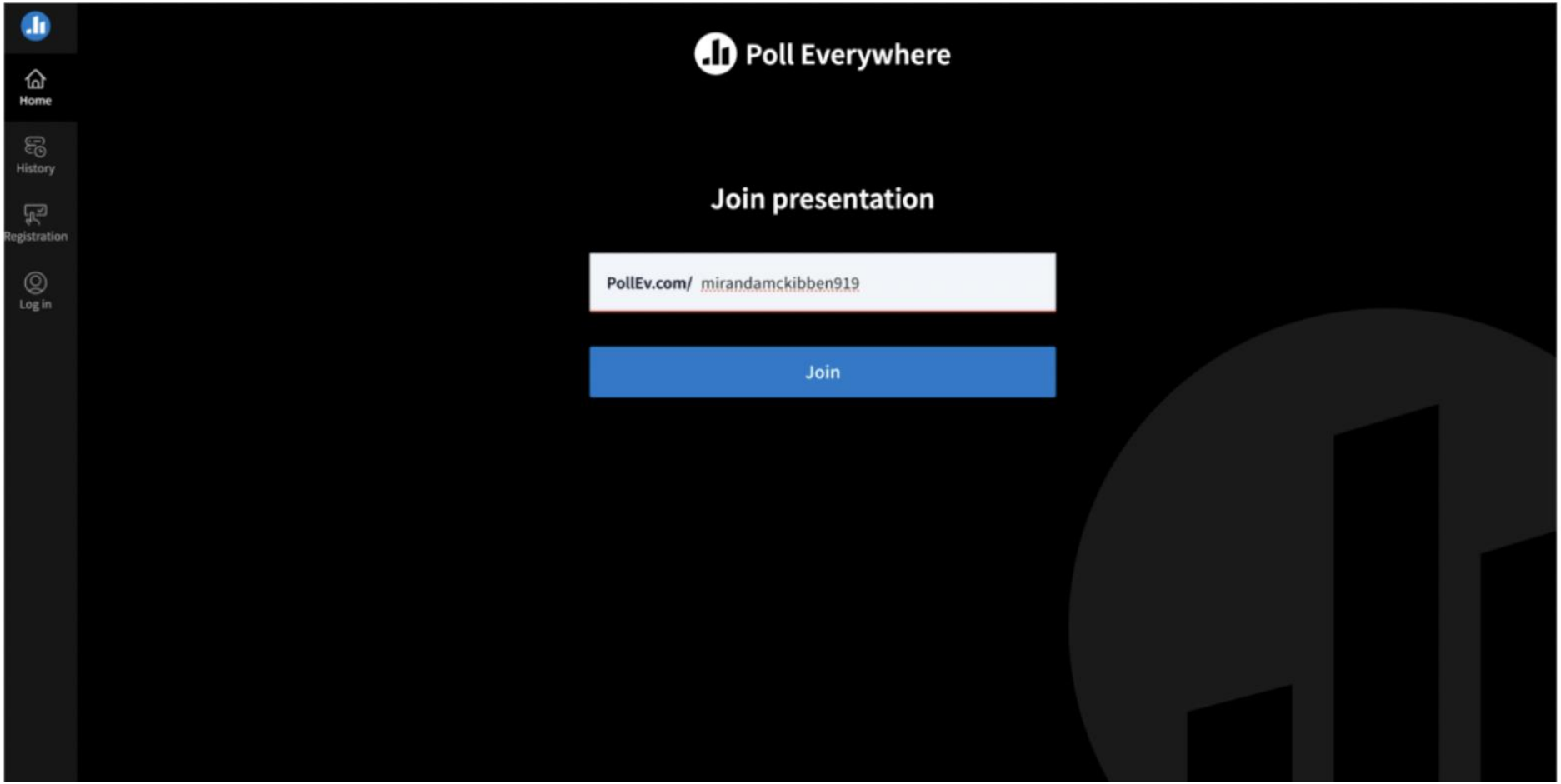
mmckibben273 Add files via upload 97935e3 · 26 minutes ago History

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

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

Learning objectives 01/09/2026

- Students will be able to identify whether variables in a dataset are independent or dependent variables.
- Students will be able to identify whether a variable is qualitative/categorical or quantitative/continuous
- Students will be able to write code that produces multiple plots suited for both categorical and continuous data.

Which code below is written correctly?

7 <- x

0

x <- 5

0

Data <- (4,6,8,7)

0

5 <- 2^2

0

None of the above

0

The first step in graphing is identifying your independent and dependent variables

Which type of variable changes in response to the other

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 (categorical) or quantitative (continuous/discrete)?

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 do trees in a forest differ in height?

Step 2: Are your variables qualitative (categorical) or quantitative (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 dependent variable continuous or categorical?

Categorical

0

Continuous

0

You want to measure how the average temperature changes over the course of 7 months. What kind of variable is your independent variable?

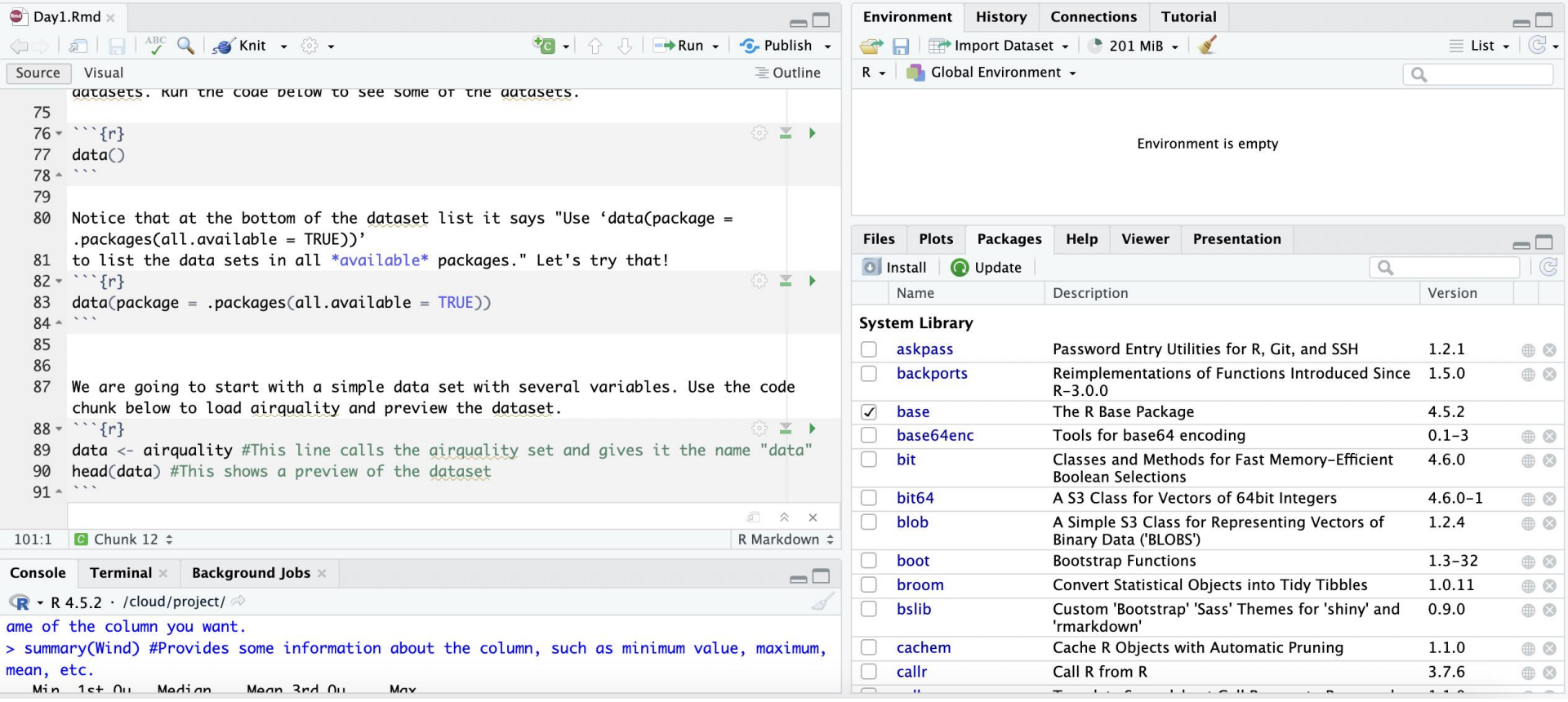
Types of plots to use for different types of data

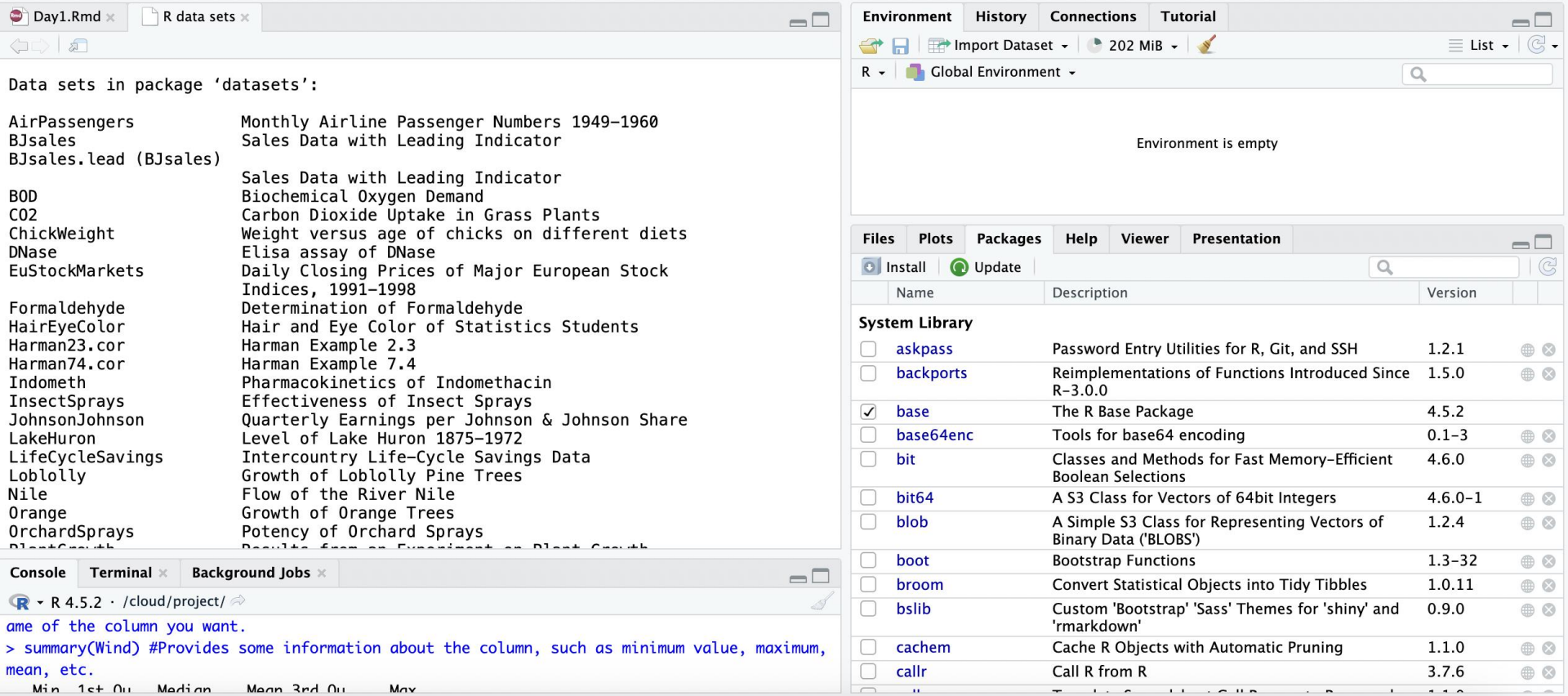
Categorical

Bar chart
Pie chart
Box and Whisker Plot

Continuous

Line graph
Scatter Plot
Histogram


Day1.Rmd
Knit
Run
Publish
Source
Visual
Outline
datasets. Run the code below to see some of the datasets.
```{r}
data()
```
Notice that at the bottom of the dataset list it says "Use 'data(package = .packages(all.available = TRUE))'
to list the data sets in all *available* packages." Let's try that!
```{r}
data(package = .packages(all.available = TRUE))
```
We are going to start with a simple data set with several variables. Use the code chunk below to load airquality and preview the dataset.
```{r}
data <- airquality #This line calls the airquality set and gives it the name "data"
head(data) #This shows a preview of the dataset
```
101:1
Chunk 12
R Markdown
Console
Terminal
Background Jobs
R 4.5.2 · /cloud/project/
ame of the column you want.
> summary(Wind) #Provides some information about the column, such as minimum value, maximum, mean, etc.
Min. 1st Qu. Median Mean 3rd Qu. Max.
Environment
History
Connections
Tutorial
Import Dataset
201 MiB
List
R
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
Presentation
Install
Update
Name
Description
Version
System Library
askpass
Password Entry Utilities for R, Git, and SSH
1.2.1
backports
Reimplementations of Functions Introduced Since R-3.0.0
1.5.0
base
The R Base Package
4.5.2
base64enc
Tools for base64 encoding
0.1-3
bit
Classes and Methods for Fast Memory-Efficient Boolean Selections
4.6.0
bit64
A S3 Class for Vectors of 64bit Integers
4.6.0-1
blob
A Simple S3 Class for Representing Vectors of Binary Data ('BLOBS')
1.2.4
boot
Bootstrap Functions
1.3-32
broom
Convert Statistical Objects into Tidy Tibbles
1.0.11
bslib
Custom 'Bootstrap' 'Sass' Themes for 'shiny' and 'rmarkdown'
0.9.0
cachem
Cache R Objects with Automatic Pruning
1.1.0
callr
Call R from R
3.7.6


Day1.Rmd
R data sets
Data sets in package 'datasets':
AirPassengers Monthly Airline Passenger Numbers 1949-1960
BJsales Sales Data with Leading Indicator
BJsales.lead (BJsales)
Sales Data with Leading Indicator
BOD Biochemical Oxygen Demand
CO2 Carbon Dioxide Uptake in Grass Plants
ChickWeight Weight versus age of chicks on different diets
DNase Elisa assay of DNase
EuStockMarkets Daily Closing Prices of Major European Stock
Indices, 1991-1998
Formaldehyde Determination of Formaldehyde
HairEyeColor Hair and Eye Color of Statistics Students
Harman23.cor Harman Example 2.3
Harman74.cor Harman Example 7.4
Indometh Pharmacokinetics of Indomethacin
InsectSprays Effectiveness of Insect Sprays
JohnsonJohnson Quarterly Earnings per Johnson & Johnson Share
LakeHuron Level of Lake Huron 1875-1972
LifeCycleSavings Intercountry Life-Cycle Savings Data
Loblolly Growth of Loblolly Pine Trees
Nile Flow of the River Nile
Orange Growth of Orange Trees
OrchardSprays Potency of Orchard Sprays
Console
Terminal
Background Jobs
R 4.5.2 · /cloud/project/
ame of the column you want.
> summary(Wind) #Provides some information about the column, such as minimum value, maximum, mean, etc.
Min. 1st Qu. Median Mean 3rd Qu. Max.
Environment
History
Connections
Tutorial
Import Dataset
202 MiB
List
R
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
Presentation
Install
Update
Name
Description
Version
System Library
askpass Password Entry Utilities for R, Git, and SSH 1.2.1
backports Reimplementations of Functions Introduced Since R-3.0.0 1.5.0
base The R Base Package 4.5.2
base64enc Tools for base64 encoding 0.1-3
bit Classes and Methods for Fast Memory-Efficient Boolean Selections 4.6.0
bit64 A S3 Class for Vectors of 64bit Integers 4.6.0-1
blob A Simple S3 Class for Representing Vectors of Binary Data ('BLOBS') 1.2.4
boot Bootstrap Functions 1.3-32
broom Convert Statistical Objects into Tidy Tibbles 1.0.11
bslib Custom 'Bootstrap' 'Sass' Themes for 'shiny' and 'rmarkdown' 0.9.0
cachem Cache R Objects with Automatic Pruning 1.1.0
callr Call R from R 3.7.6

Day1.Rmd

Knit | Run | Publish

Source | Visual | Outline

```
86
87  We are going to start with a simple data set with several variables. Use the code
    chunk below to load airquality and preview the dataset.
88  ```{r}
89  data <- airquality #This line calls the airquality set and gives it the name "data"
90  head(data) #This shows a preview of the dataset
91  ```
```

	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

101:1 | Chunk 12 | R Markdown

Console | Terminal | Background Jobs

R 4.5.2 · /cloud/project/

```
ame of the column you want.
> summary(Wind) #Provides some information about the column, such as minimum value, maximum,
mean, etc.
```

Environment | History | Connections | Tutorial

Import Dataset | 202 MiB | List

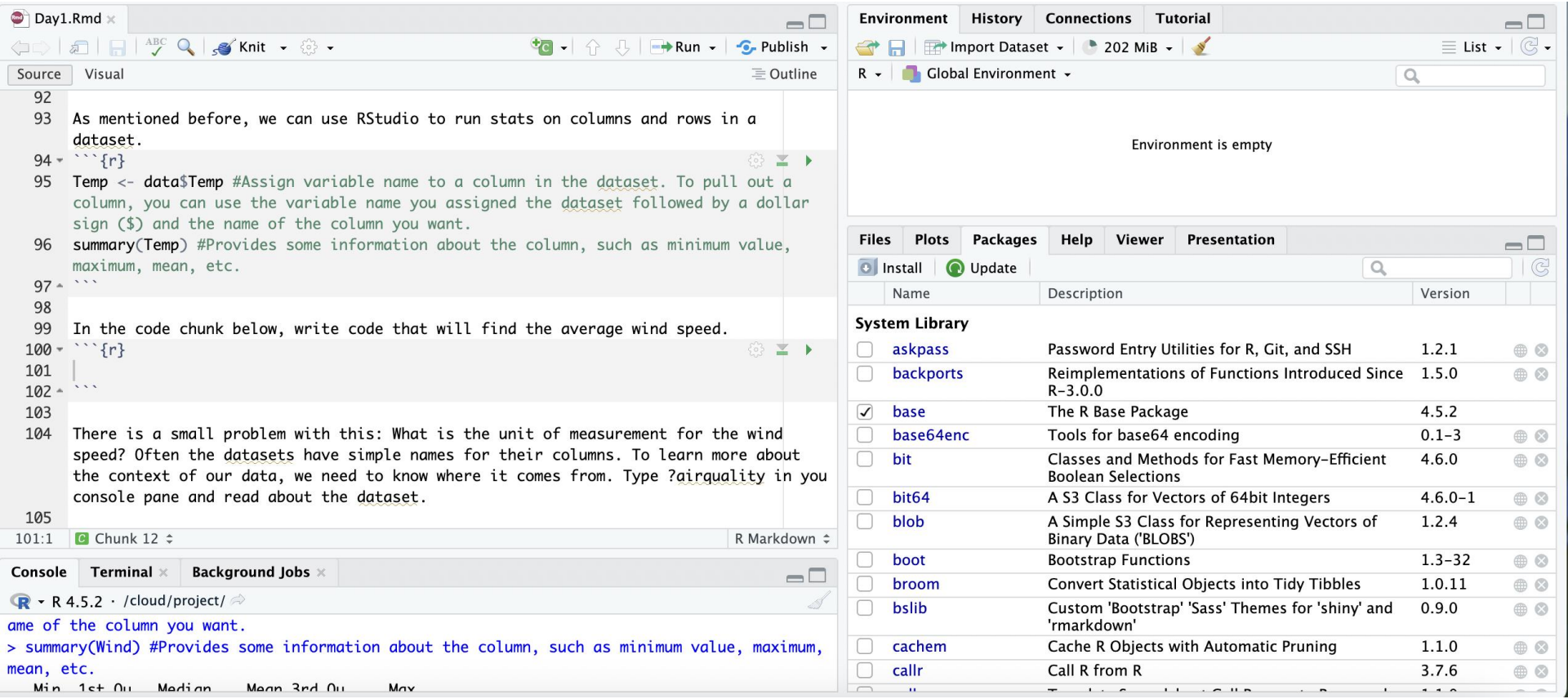
R | Global Environment

Environment is empty

Files | Plots | Packages | Help | Viewer | Presentation

Install | Update

	Name	Description	Version
System Library			
☐	askpass	Password Entry Utilities for R, Git, and SSH	1.2.1
☐	backports	Reimplementations of Functions Introduced Since R-3.0.0	1.5.0
☑	base	The R Base Package	4.5.2
☐	base64enc	Tools for base64 encoding	0.1-3
☐	bit	Classes and Methods for Fast Memory-Efficient Boolean Selections	4.6.0
☐	bit64	A S3 Class for Vectors of 64bit Integers	4.6.0-1
☐	blob	A Simple S3 Class for Representing Vectors of Binary Data ('BLOBS')	1.2.4
☐	boot	Bootstrap Functions	1.3-32
☐	broom	Convert Statistical Objects into Tidy Tibbles	1.0.11
☐	bslib	Custom 'Bootstrap' 'Sass' Themes for 'shiny' and 'rmarkdown'	0.9.0
☐	cachem	Cache R Objects with Automatic Pruning	1.1.0
☐	callr	Call R from R	3.7.6


Day1.Rmd
Knit
Run
Publish
Source
Visual
Outline
As mentioned before, we can use RStudio to run stats on columns and rows in a dataset.
```{r}
Temp <- data$Temp #Assign variable name to a column in the dataset. To pull out a column, you can use the variable name you assigned the dataset followed by a dollar sign ($) and the name of the column you want.
summary(Temp) #Provides some information about the column, such as minimum value, maximum, mean, etc.
```
In the code chunk below, write code that will find the average wind speed.
```{r}
```
There is a small problem with this: What is the unit of measurement for the wind speed? Often the datasets have simple names for their columns. To learn more about the context of our data, we need to know where it comes from. Type ?airquality in you console pane and read about the dataset.
101:1
Chunk 12
R Markdown
Console
Terminal
Background Jobs
R 4.5.2 · /cloud/project/
ame of the column you want.
> summary(Wind) #Provides some information about the column, such as minimum value, maximum, mean, etc.
Environment
History
Connections
Tutorial
Import Dataset
202 MiB
List
R
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
Presentation
Install
Update
Name
Description
Version
System Library
askpass
Password Entry Utilities for R, Git, and SSH
1.2.1
backports
Reimplementations of Functions Introduced Since R-3.0.0
1.5.0
base
The R Base Package
4.5.2
base64enc
Tools for base64 encoding
0.1-3
bit
Classes and Methods for Fast Memory-Efficient Boolean Selections
4.6.0
bit64
A S3 Class for Vectors of 64bit Integers
4.6.0-1
blob
A Simple S3 Class for Representing Vectors of Binary Data ('BLOBS')
1.2.4
boot
Bootstrap Functions
1.3-32
broom
Convert Statistical Objects into Tidy Tibbles
1.0.11
bslib
Custom 'Bootstrap' 'Sass' Themes for 'shiny' and 'rmarkdown'
0.9.0
cachem
Cache R Objects with Automatic Pruning
1.1.0
callr
Call R from R
3.7.6

What is the average wind speed in this dataset? If you have time, use ?airquality to find the unit of measurement, as described beneath the code chunk.

Nobody has responded yet.

Hang tight! Responses are coming in.

Day1.Rmd

Knit | Run | Publish

Source | Visual | Outline

```
maximum, mean, etc.
```

```

In the code chunk below, write code that will find the average wind speed.
```{r}
Wind <- data$Wind #Assign variable name to a column in the dataset. To pull out a column, you can use the variable name you assigned the dataset followed by a dollar sign ($) and the name of the column you want.
summary(Wind) #Provides some information about the column, such as minimum value, maximum, mean, etc.

mean(Wind)
```
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.700   7.400   9.700   9.958  11.500  20.700
[1] 9.957516
```

Show in New Window

```

There is a small problem with this: What is the unit of measurement for the wind speed? Often the datasets have simple names for their columns. To learn more about
```

104:10 | Chunk 12 | R Markdown

Console | Terminal | Background Jobs

R 4.5.2 · /cloud/project/

```
> mean(Wind)
[1] 9.957516
>
```

Environment | History | Connections | Tutorial

Import Dataset | 292 MiB | List

R | Global Environment

Data

data	153 obs. of 6 variables

Values

Wind	num [1:153] 7.4 8 12.6 11.5 14.3 14.9 8.6 13.8 20.1 …

Files | Plots | Packages | Help | Viewer | Presentation

Install | Update

	Name	Description	Version
System Library			
☐	askpass	Password Entry Utilities for R, Git, and SSH	1.2.1
☐	backports	Reimplementations of Functions Introduced Since R-3.0.0	1.5.0
☐	base	The R Base Package	4.5.2
☐	base64enc	Tools for base64 encoding	0.1-3
☐	bit	Classes and Methods for Fast Memory-Efficient Boolean Selections	4.6.0
☐	bit64	A S3 Class for Vectors of 64bit Integers	4.6.0-1
☐	blob	A Simple S3 Class for Representing Vectors of Binary Data ('BLOBS')	1.2.4
☐	boot	Bootstrap Functions	1.3-32
☐	broom	Convert Statistical Objects into Tidy Tibbles	1.0.11
☐	bslib	Custom 'Bootstrap' 'Sass' Themes for 'shiny' and 'rmarkdown'	0.9.0
☐	cachem	Cache R Objects with Automatic Pruning	1.1.0
☐	callr	Call R from R	3.7.6

Find the distribution of temperatures in the month of May in the code chunk below:

```{r}
summary(May)
```

```
     Ozone           Solar.R          Wind             Temp           Month       Day       
 Min.   :  1.00   Min.   :  8.0   Min.   : 5.70   Min.   :56.00   Min.   :5   Min.   : 1.0  
 1st Qu.: 11.00   1st Qu.: 72.0   1st Qu.: 8.90   1st Qu.:60.00   1st Qu.:5   1st Qu.: 8.5  
 Median : 18.00   Median :194.0   Median :11.50   Median :66.00   Median :5   Median :16.0  
 Mean   : 23.62   Mean   :181.3   Mean   :11.62   Mean   :65.55   Mean   :5   Mean   :16.0  
 3rd Qu.: 31.50   3rd Qu.:284.5   3rd Qu.:14.05   3rd Qu.:69.00   3rd Qu.:5   3rd Qu.:23.5  
 Max.   :115.00   Max.   :334.0   Max.   :20.10   Max.   :81.00   Max.   :5   Max.   :31.0  
 NA's   :5        NA's   :4                                                                 
```

Find the distribution of temperatures in the month of May in the code chunk below:

```{r}
summary(May)
```

```
     Ozone           Solar.R          Wind             Temp           Month       Day       
 Min.   :  1.00   Min.   :  8.0   Min.   : 5.70   Min.   :56.00   Min.   :5   Min.   : 1.0  
 1st Qu.: 11.00   1st Qu.: 72.0   1st Qu.: 8.90   1st Qu.:60.00   1st Qu.:5   1st Qu.: 8.5  
 Median : 18.00   Median :194.0   Median :11.50   Median :66.00   Median :5   Median :16.0  
 Mean   : 23.62   Mean   :181.3   Mean   :11.62   Mean   :65.55   Mean   :5   Mean   :16.0  
 3rd Qu.: 31.50   3rd Qu.:284.5   3rd Qu.:14.05   3rd Qu.:69.00   3rd Qu.:5   3rd Qu.:23.5  
 Max.   :115.00   Max.   :334.0   Max.   :20.10   Max.   :81.00   Max.   :5   Max.   :31.0  
 NA's   :5        NA's   :4
```

Find the distribution of temperatures in August:

```{r}
August <- filter(data, Month == 8)
summary(August)
```

```
     Ozone           Solar.R          Wind             Temp           Month       Day       
 Min.   :  9.00   Min.   : 24.0   Min.   : 2.300   Min.   :72.00   Min.   :8   Min.   : 1.0  
 1st Qu.: 28.75   1st Qu.:107.0   1st Qu.: 6.600   1st Qu.:79.00   1st Qu.:8   1st Qu.: 8.5  
 Median : 52.00   Median :197.5   Median : 8.600   Median :82.00   Median :8   Median :16.0  
 Mean   : 59.96   Mean   :171.9   Mean   : 8.794   Mean   :83.97   Mean   :8   Mean   :16.0  
 3rd Qu.: 82.50   3rd Qu.:231.0   3rd Qu.:11.200   3rd Qu.:88.50   3rd Qu.:8   3rd Qu.:23.5  
 Max.   :168.00   Max.   :273.0   Max.   :15.500   Max.   :97.00   Max.   :8   Max.   :31.0  
 NA's   :5        NA's   :3
```

Day1.Rmd

Knit | Run | Publish

Source | Visual | Outline

```
106
107 There is a small problem with this: What is the unit of measurement for the wind
    speed? Often the datasets have simple names for their columns. To learn more about
    the context of our data, we need to know where it comes from. Type ?airquality in you
    console pane and read about the dataset.
108
109
110 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.
111 ```{r}
112 plot(data$Month,data$Temp)
113 ```
114
115 This plot is alright, but can be improved. Do you notice what is missing?
116
117 In addition to adding/changing labels on a plot, we can specify in our code how we
    want to visualize the data. When we did a simple plot command RStudio automatically
    created a scatterplot because of the input format. Let's try another plot, like a
    boxplot.
118 ```{r}
```

104:10 | Chunk 12 | R Markdown

Console | Terminal | Background Jobs

R 4.5.2 · /cloud/project/

```
> mean(Wind)
[1] 9.957516
>
```

Environment | History | Connections | Tutorial

Import Dataset | 293 MiB | List

R | Global Environment

Data

data	153 obs. of 6 variables

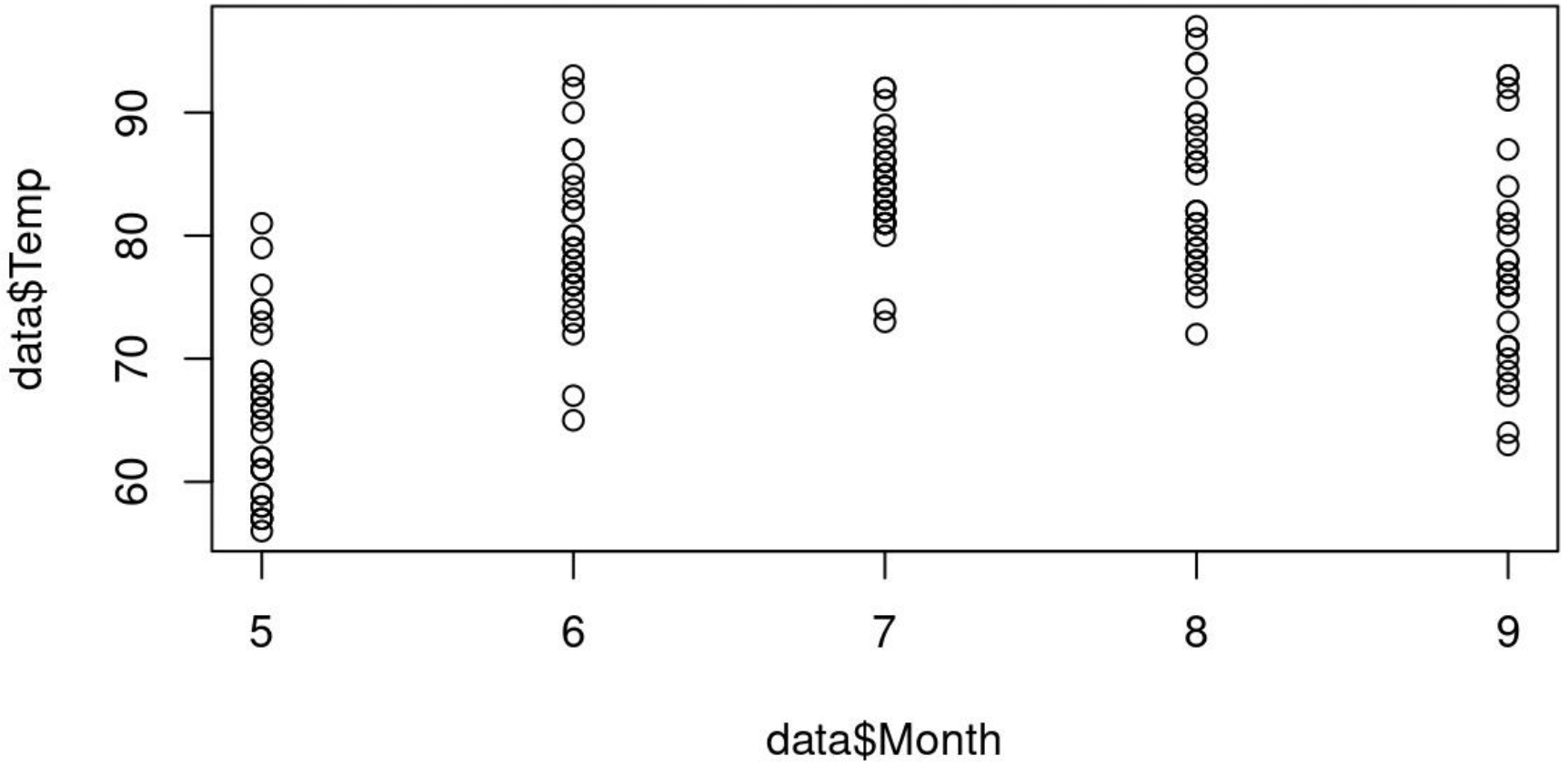
Values

Wind	num [1:153] 7.4 8 12.6 11.5 14.3 14.9 8.6 13.8 20.1 …

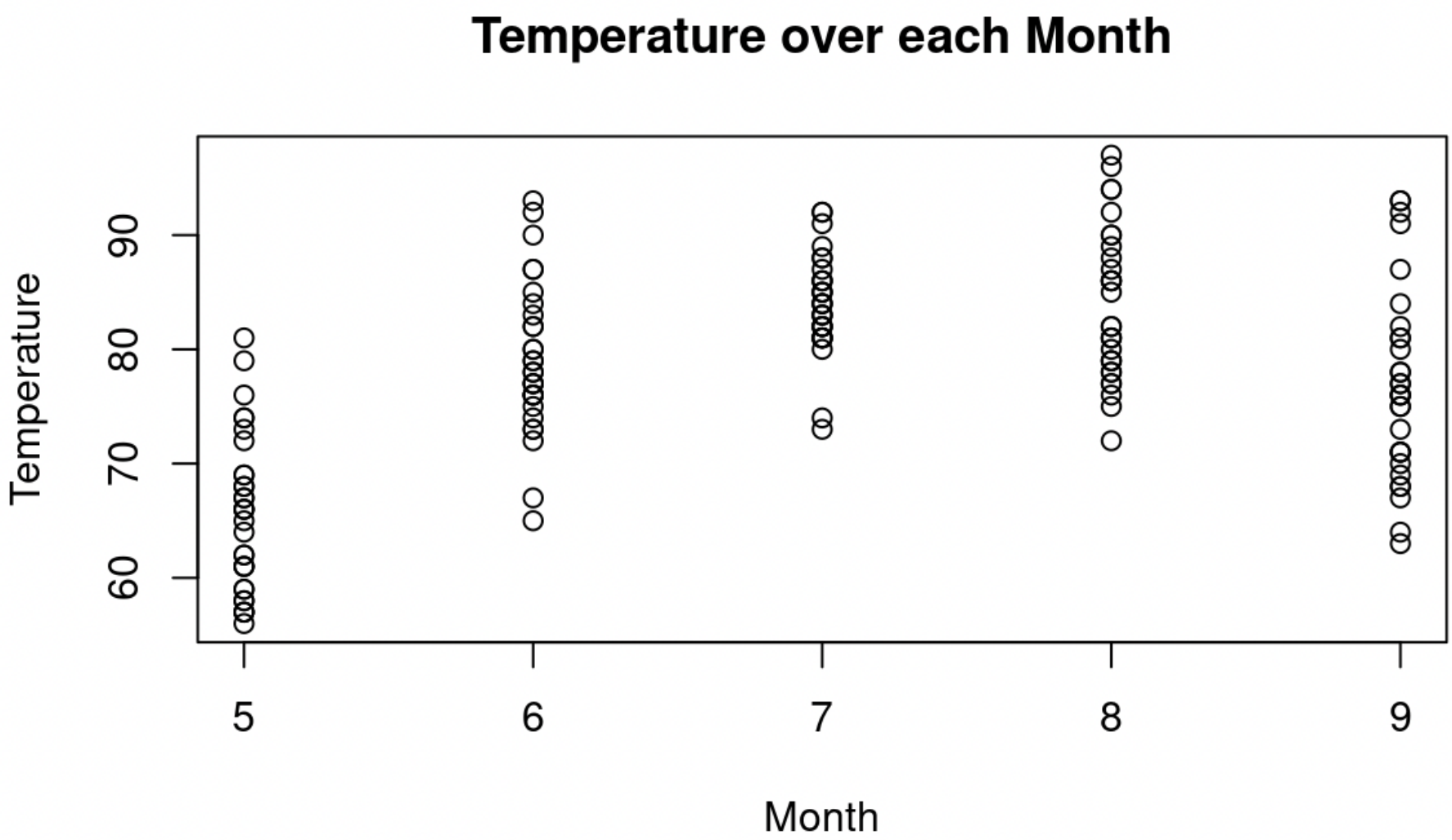
Files | Plots | Packages | Help | Viewer | Presentation

Install | Update

	Name	Description	Version
System Library			
☐	askpass	Password Entry Utilities for R, Git, and SSH	1.2.1
☐	backports	Reimplementations of Functions Introduced Since R-3.0.0	1.5.0
☑	base	The R Base Package	4.5.2
☐	base64enc	Tools for base64 encoding	0.1-3
☐	bit	Classes and Methods for Fast Memory-Efficient Boolean Selections	4.6.0
☐	bit64	A S3 Class for Vectors of 64bit Integers	4.6.0-1
☐	blob	A Simple S3 Class for Representing Vectors of Binary Data ('BLOBS')	1.2.4
☐	boot	Bootstrap Functions	1.3-32
☐	broom	Convert Statistical Objects into Tidy Tibbles	1.0.11
☐	bslib	Custom 'Bootstrap' 'Sass' Themes for 'shiny' and 'rmarkdown'	0.9.0
☐	cachem	Cache R Objects with Automatic Pruning	1.1.0
☐	callr	Call R from R	3.7.6



What kind of variable is “month”? What kind of variables do we plot with scatterplots?


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

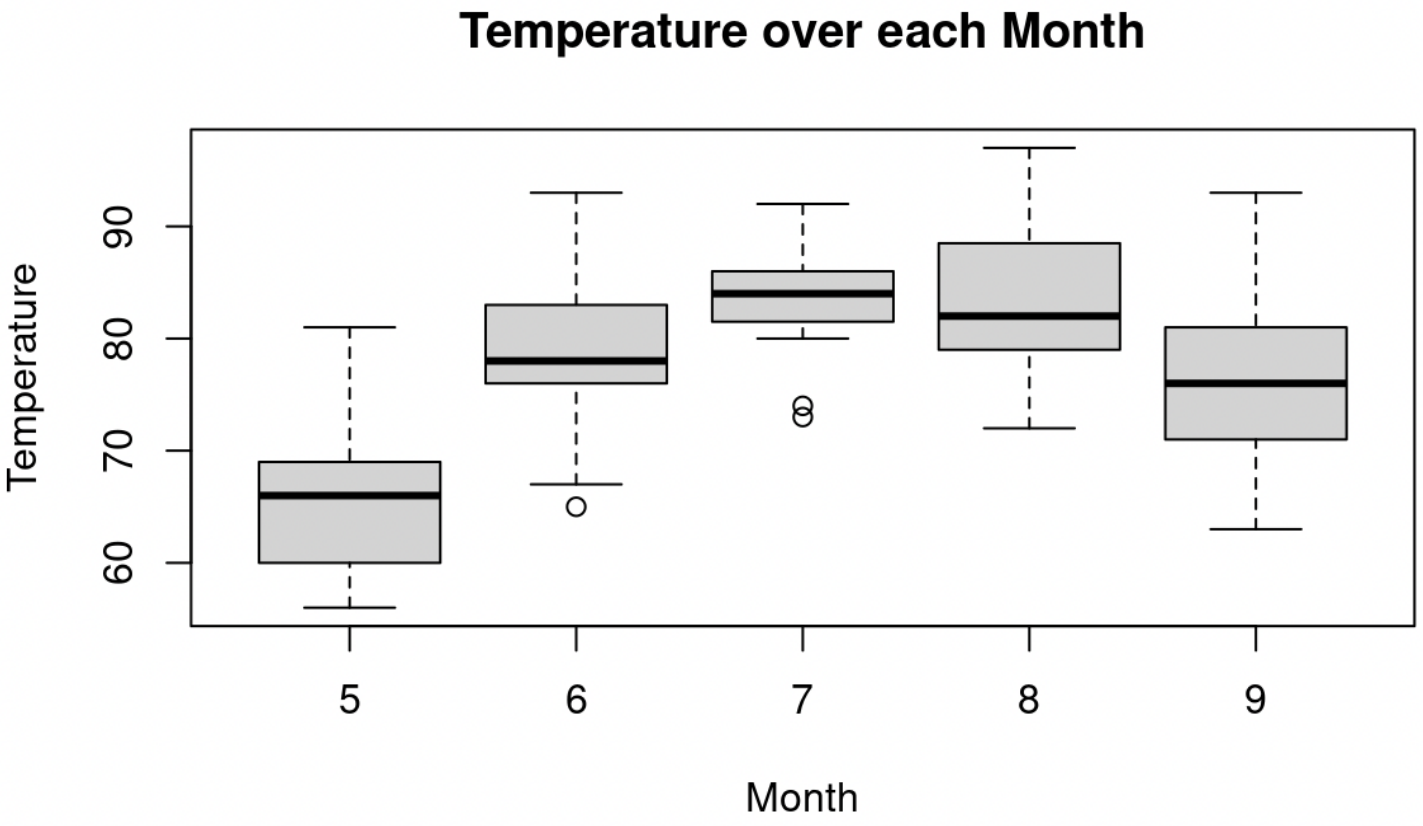
In addition to adding/changing labels on a plot, we can specify in our code how we want to visualize the data. When we did a simple plot command RStudio automatically created a scatterplot because of the input format. Let's try another plot, like a boxplot.

```{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}

```


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

Have you made a box plot?

Yes

0%

No

0%

````
Now make your own boxplot for wind speed each month.
```{r}
plot(data$Month,data$Wind, xlab = "Month", ylab = "Speed",main = "Wind speed over each Month")
boxplot(Wind~Month, data=data, xlab = "Month", ylab = "Speed",main = "Wind speed over each Month") #Notice the input for a boxplot is dependent~independent variable
```

The boxplot is nice, but with other packages we can make it even better.
RStudio has a package called ggplot2 (Grammar of Graphics Plot).

```{r}
library(tidyverse) #RStudio has packages that contain lots of premade functions that can help us withour analysis. To load a package, type library() and put the name of the package in parentheses
library(ggplot2)
data$Month <- as.factor(data$Month) #To color code the months, group the data points that belong to each month first.
ggplot(data = data)+ #Specify data set
  aes(x = Month, y = Temp, fill = Month)+ #Your aesthetics. Identify your x and y
````

125:81 Chunk 15 R Markdown

Console Terminal Background Jobs

R 4.5.2 · /cloud/project/

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

Data

data	153 obs. of 6 variables

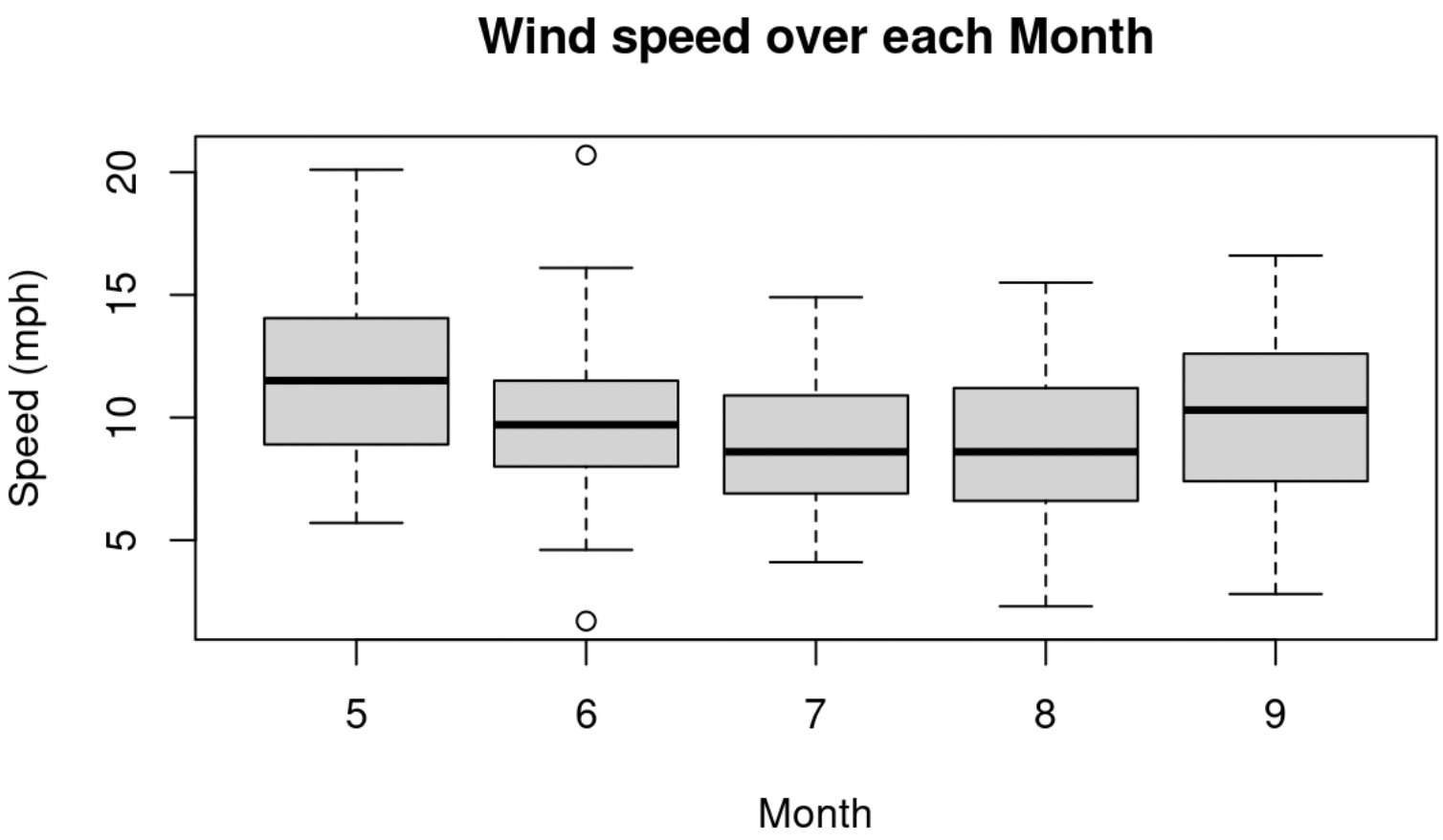
Values

Wind	num [1:153] 7.4 8 12.6 11.5 14.3 14.9 8.6 13.8 20.1 …

Files Plots Packages Help Viewer Presentation

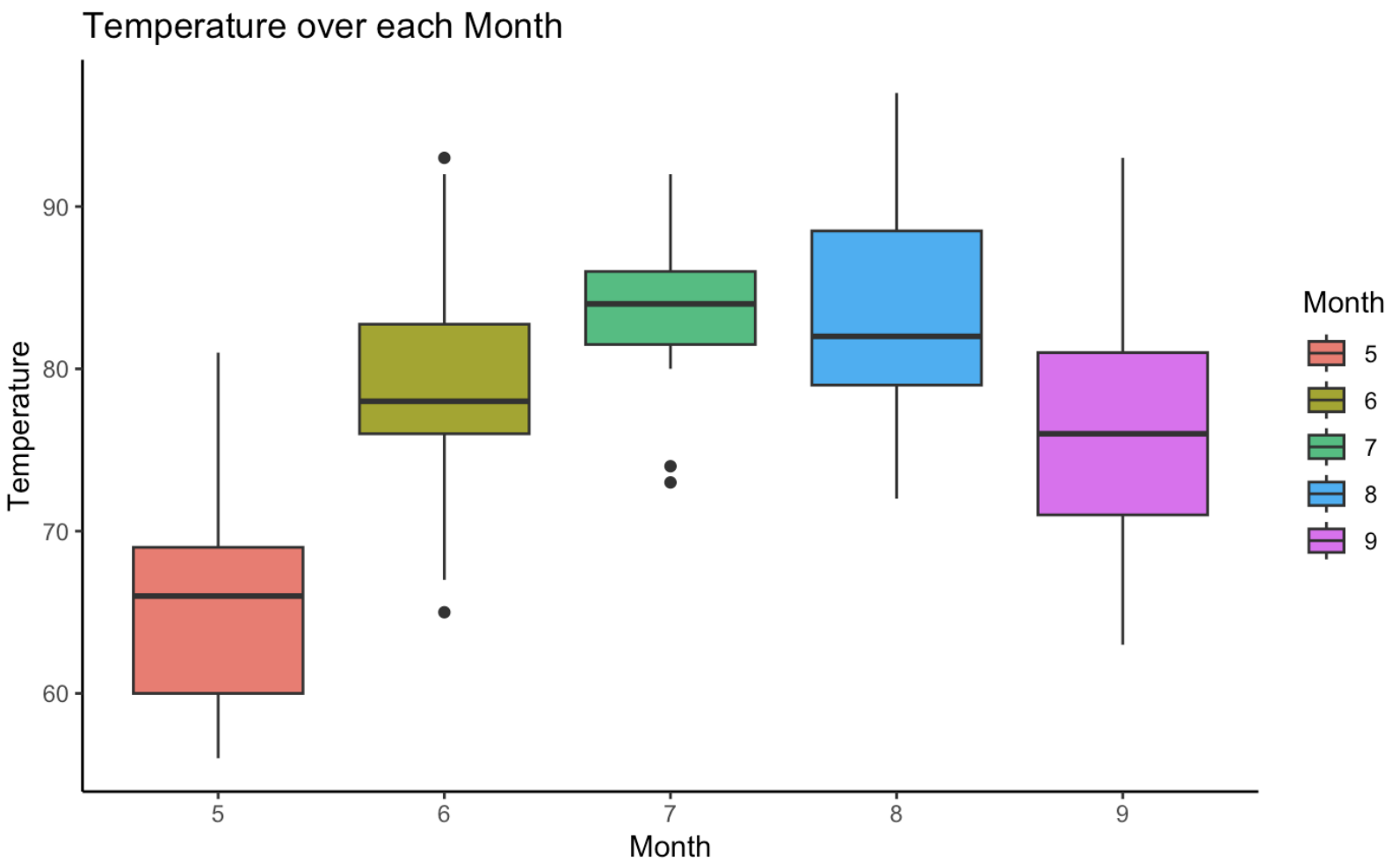
Install Update

	Name	Description	Version
System Library			
☐	askpass	Password Entry Utilities for R, Git, and SSH	1.2.1
☐	backports	Reimplementations of Functions Introduced Since R-3.0.0	1.5.0
☑	base	The R Base Package	4.5.2
☐	base64enc	Tools for base64 encoding	0.1-3
☐	bit	Classes and Methods for Fast Memory-Efficient Boolean Selections	4.6.0
☐	bit64	A S3 Class for Vectors of 64bit Integers	4.6.0-1
☐	blob	A Simple S3 Class for Representing Vectors of Binary Data ('BLOBS')	1.2.4
☐	boot	Bootstrap Functions	1.3-32
☐	broom	Convert Statistical Objects into Tidy Tibbles	1.0.11
☐	bslib	Custom 'Bootstrap' 'Sass' Themes for 'shiny' and 'rmarkdown'	0.9.0
☐	cachem	Cache R Objects with Automatic Pruning	1.1.0
☐	callr	Call R from R	3.7.6
```


Wind speed over each Month
Speed (mph)
5
10
15
20
5
6
7
8
9
Month

The boxplot is nice, but with other packages we can make it even better.
RStudio has a package called ggplot2 (Grammar of Graphics Plot).

```{r}
library(tidyverse) #RStudio has packages that contain lots of premade functions that can help us withour analysis. To load a package, type library() and put the name of the package in parentheses
library(ggplot2)
data$Month <- as.factor(data$Month) #To color code the months, group the data points that belong to each month first.
ggplot(data = data)+ #Specify data set
  aes(x = Month, y = Temp, fill = Month)+ #Your aesthetics. Identify your x and y axis as well as how you want to color code.
  geom_boxplot()+ #What type of figure/plot to create
  ylab("Temperature") + # x label
  xlab("Month") + # y label
  ggtitle("Temperature over each Month")+ theme_classic()
```


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

```{r }
library(palmerpenguins) #Load the package containing the palmer penguin dataset
data1 <- palmerpenguins::penguins #load the specific penguin dataset from the package

ggplot(data1) + #This line tells RStudio what dataset you are pulling data from
  aes(x = island, fill = species)+ #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)
```

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}

```

Have you made a bar graph?

Yes

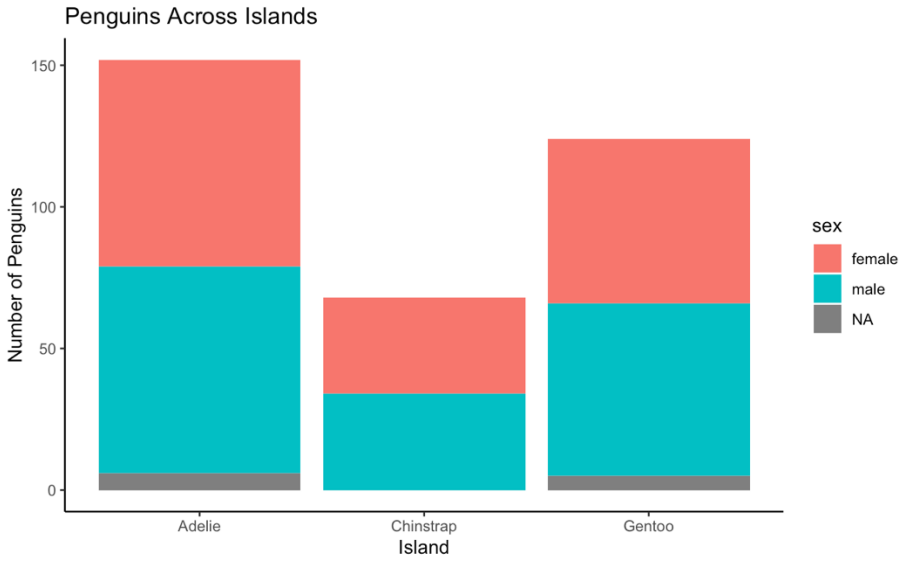
0%

No

0%

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)
```



```{r}
mytable <- table(data1$species) #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 Species\n (with sample sizes)")
```

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

```{r}

```

Have you made a pie chart?

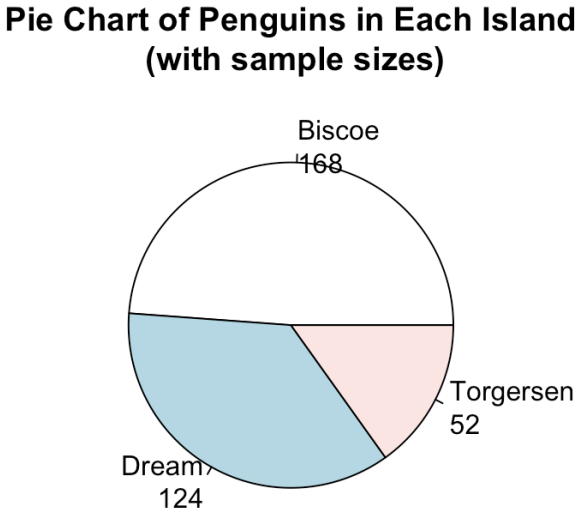
Yes

0%

No

0%

```{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)")
```

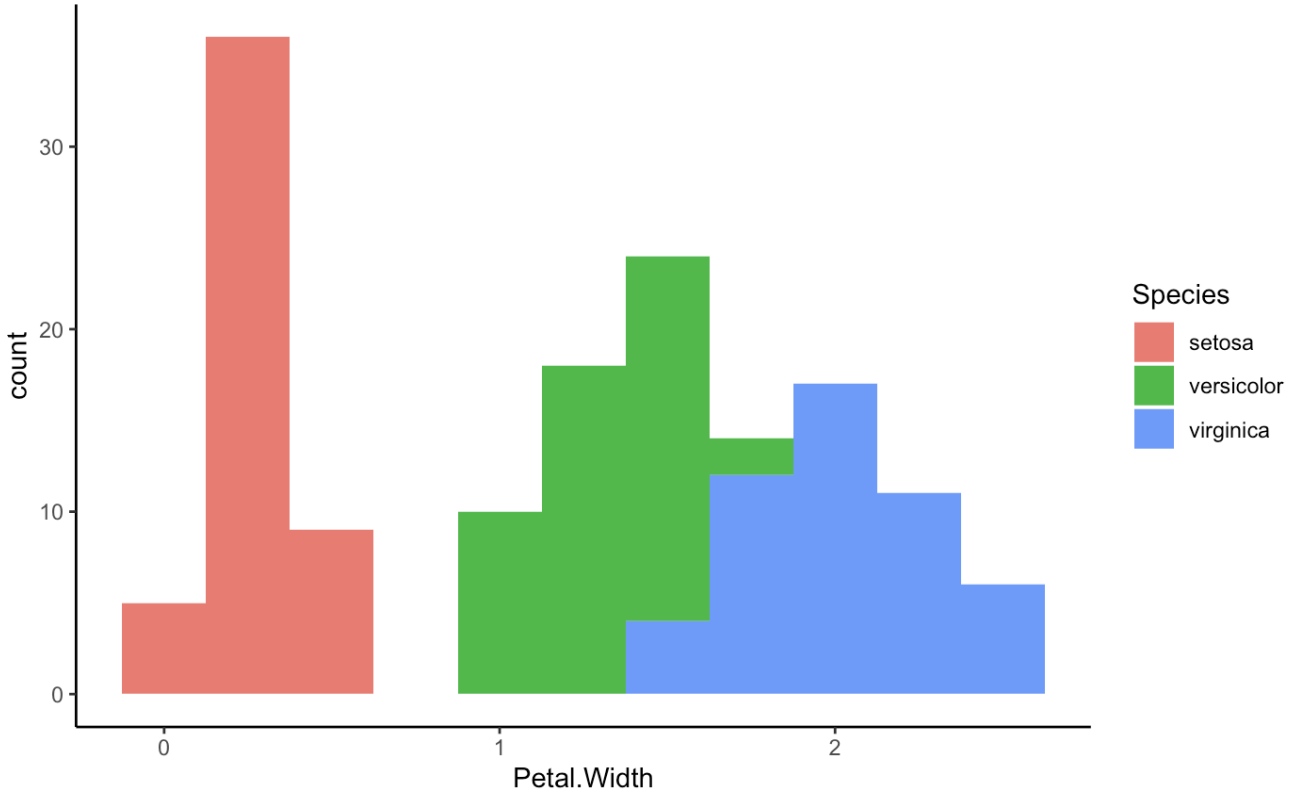


Create your own histogram looking at petal width across species.

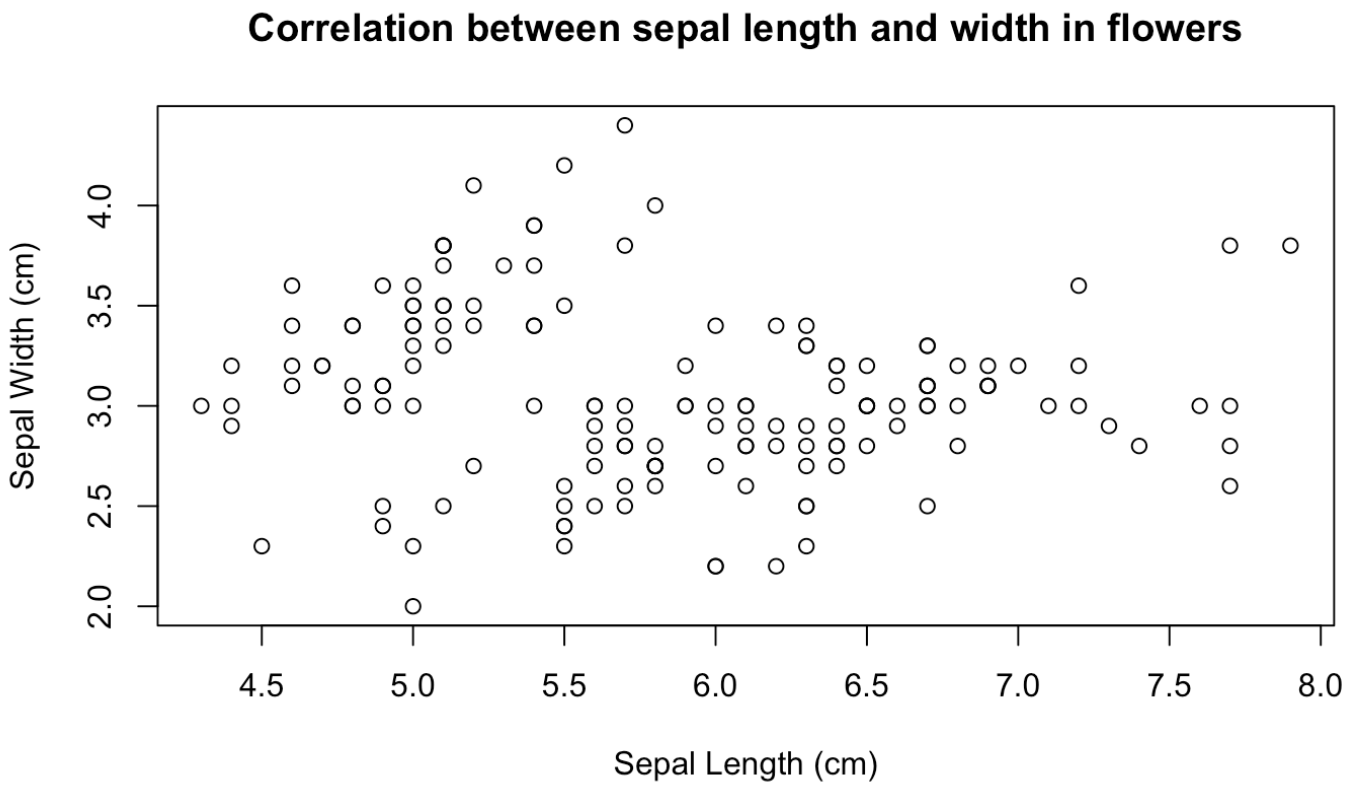
```{r}

```

Have you made a histogram?

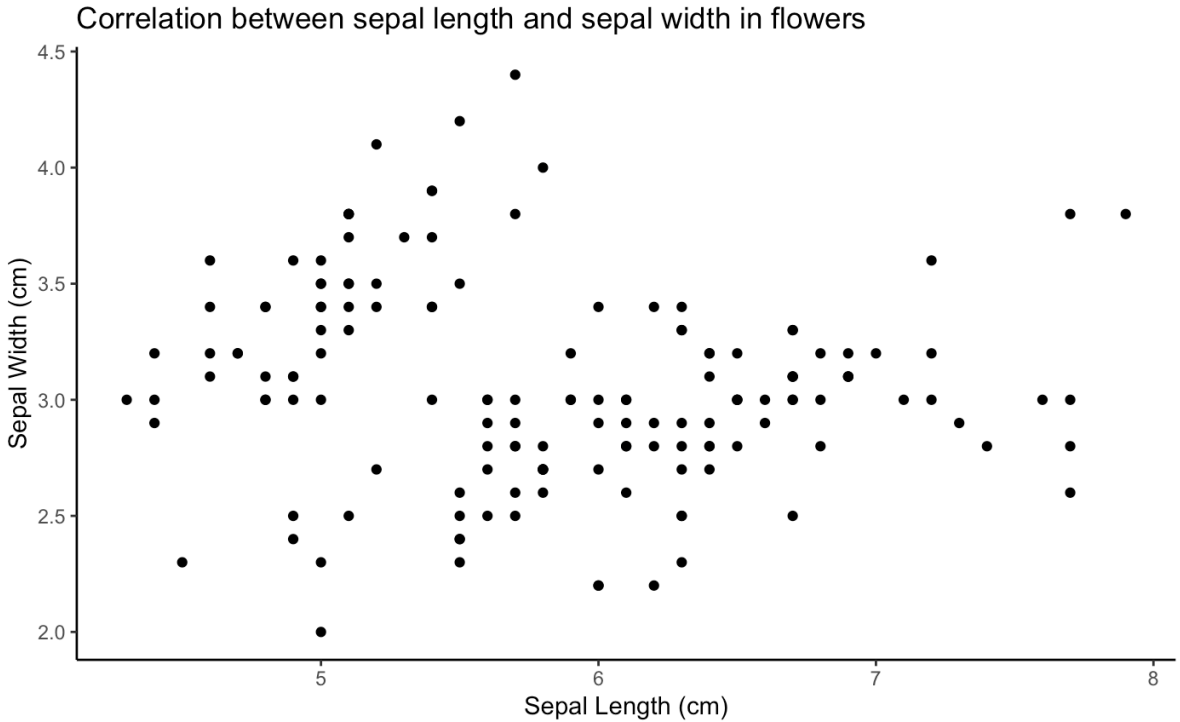
```{r}
ggplot(data2, aes(x = Petal.Width, fill=Species)) +
  geom_histogram(binwidth = 0.25)+ #Try changing the bin width and rerunning the code to see how it changes your visualization. When deciding binwidths for your histograms, make sure to think about the range/length of your x axis.
  theme_classic()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



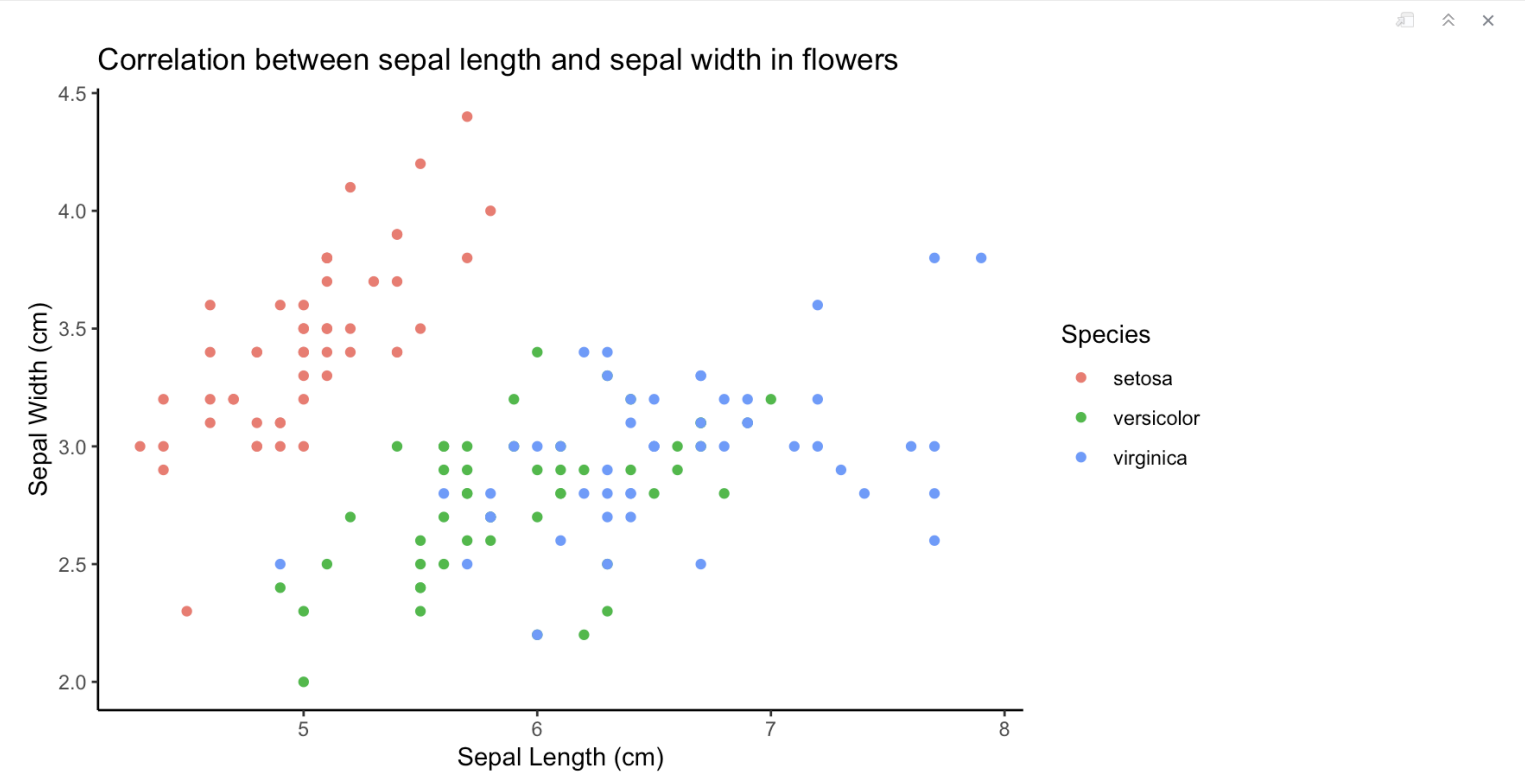
```{r}
plot(data2$Sepal.Length, data2$Sepal.Width,xlab = "Sepal Length (cm)", ylab = "Sepal Width (cm)",main = "Correlation between sepal length and width in flowers")
```



```{r}
ggplot(data = data2) +
  aes(x = Sepal.Length, y = Sepal.Width) +
  geom_point() +
  theme_classic() +
  ylab("Sepal Width (cm)")+
  xlab("Sepal Length (cm)")+
  ggtitle("Correlation between sepal length and sepal width in flowers")

```



```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()+
  ylab("Sepal Width (cm)")+
  xlab("Sepal Length (cm)")+
  ggtitle("Correlation between sepal length and sepal width in flowers")
```



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.