The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	000000	00000	0000000	0000	000

## Introduction to the Grammar of Graphics II

Prof. Wells

STA 209, 2/1/23

The ggp1ot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	000000	00000	0000000	0000	000
Outline					

In this lecture, we will...

The ggplot2 Package 000	Scatterplots 000000	Line Graphs 00000	Histograms 0000000	Boxplots 0000	Bar Charts 000

Outline

In this lecture, we will...

- Discuss the Grammar of Graphics
- Decompose particular graphics using the GG paradigm
- Introduce the ggplot2 package for R graphics

The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
●OO	000000	00000	0000000	0000	000

## Section 1

# The ggplot2 Package

The ggplot2 Package O●O	Scatterplots 000000	Line Graphs 00000	Histograms 0000000	Boxplots 0000	Bar Charts 000
The ggplot2 s	syntax				

• We will use the ggplot function in the ggplot2 package for data visualization in accordance with the grammar of graphics.

The ggp1ot2 Package O●O	Scatterplots 000000	Line Graphs 00000	Histograms 0000000	Boxplots 0000	Bar Charts 000
The ggplot2 syr	itax				

- We will use the ggplot function in the ggplot2 package for data visualization in accordance with the grammar of graphics.
- Recall the guiding principle:

A statistical graphic is a mapping of data variables to aesthetic attributes of geometric objects.

The ggplot2 Package O●O	Scatterplots 000000	Line Graphs 00000	Histograms 0000000	Boxplots 0000	Bar Charts 000
The ggplot2 sy	ntax				

- We will use the ggplot function in the ggplot2 package for data visualization in accordance with the grammar of graphics.
- Recall the guiding principle:

A statistical graphic is a mapping of data variables to aesthetic attributes of geometric objects.

• The code for graphics will (almost) always take the following general form:

```
ggplot(data = ---, mapping = aes(---)) +
geom_---(---)
```

The ggplot2 Package O●O	Scatterplots 000000	Line Graphs 00000	Histograms 0000000	Boxplots 0000	Bar Charts 000
The ggplot2 syn	itax				

- We will use the ggplot function in the ggplot2 package for data visualization in accordance with the grammar of graphics.
- Recall the guiding principle:

A statistical graphic is a mapping of data variables to aesthetic attributes of geometric objects.

• The code for graphics will (almost) always take the following general form:

```
ggplot(data = ---, mapping = aes(---)) +
geom_---(---)
```

• For brevity, the above code can also be written as:

```
ggplot(---, aes(---)) +
geom_---(---)
```

• R will assume that the first argument is the data argument and the second argument is the mapping argument.

The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
OO●	000000	00000	0000000	0000	000

#### The Five Named Graphs

• We focus on just 5 graphs fundamental to statistics:

The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
OO●	000000	00000	0000000	0000	000

#### The Five Named Graphs

- We focus on just 5 graphs fundamental to statistics:
  - Scatterplots (geom\_point)
  - 2 Linegraphs (geom\_line)
  - 8 Histograms (geom\_histogram)
  - Boxplots (geom\_boxplot)
  - 6 Barplots (geom\_bar)

The ggp1ot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
00●	000000	00000	0000000	0000	000

#### The Five Named Graphs

- We focus on just 5 graphs fundamental to statistics:
  - Scatterplots (geom\_point)
  - 2 Linegraphs (geom\_line)
  - Histograms (geom\_histogram)
  - Ø Boxplots (geom\_boxplot)
  - Barplots (geom\_bar)

• We'll use a common data set to investigate each graph: the Portland Biketown data: biketown <- read\_csv("biketown.csv")

• Biketown PDX is a bike-sharing system in Portland, OR, owned by PBOT, operated by Lyft, and sponsored by Nike

The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	●00000	00000	0000000	0000	000

## Section 2

Scatterplots

The ggplot2 Package 000	Scatterplots ○●○○○○	Line Graphs 00000	Histograms 0000000	Boxplots 0000	Bar Charts 000
Creating Scatte	rplots				

We might expect a positive relationship between Duration and Distance\_Miles
 ggplot(data = biketown, mapping = aes(x = Duration, y = Distance\_Miles)) +
 geom\_point()



• But this graphic is hard to read due to large number of points plotted in close proximity.

The ggplot2 Package 000	Scatterplots 00●000	Line Graphs 00000	Histograms 0000000	Boxplots 0000	Bar Charts 000
Overplotting					

• Overplotting occurs when a large number of points are plotted in close proximity, making it difficult to accurately distinguish true number of points in a region.

The ggplot2 Package 000	Scatterplots 00●000	Line Graphs 00000	Histograms 0000000	Boxplots 0000	Bar Charts 000
Overplotting					

- Overplotting occurs when a large number of points are plotted in close proximity, making it difficult to accurately distinguish true number of points in a region.
  - Can be corrected by making points more transparent via the alpha aesthetic:

The ggp1ot2 Package 000	Scatterplots 00●000	Line Graphs 00000	Histograms 0000000	Boxplots 0000	Bar Charts 000
Overplotting					

 Overplotting occurs when a large number of points are plotted in close proximity, making it difficult to accurately distinguish true number of points in a region.

 Can be corrected by making points more transparent via the alpha aesthetic: ggplot(data = biketown, mapping = aes(x = Duration, y = Distance\_Miles)) + geom\_point(alpha = 0.15)



The ggplot2 Package 000	Scatterplots 000●00	Line Graphs 00000	Histograms 0000000	Boxplots 0000	Bar Charts 000
Overplotting II					

• We can also focus on just part of the graph by controlling the limits of the axes:

The ggplot2 Package 000	Scatterplots 000●00	Line Graphs 00000	Histograms 0000000	Boxplots 0000	Bar Charts 000
Overplotting II					

• We can also focus on just part of the graph by controlling the limits of the axes:

```
ggplot(data = biketown, mapping = aes(x = Duration, y = Distance_Miles)) +
geom_point(alpha = .15)+
scale_x_continuous(limits = c(0, 60))+
scale_y_continuous(limits = c(0, 10))
```



The ggplot2 Package 000	Scatterplots 0000●0	Line Graphs 00000	Histograms 0000000	Boxplots 0000	Bar Charts 000
Overplotting III					

• Alternatively, can manipulate data set by jittering points a small random amount so that they no longer lie on top of each other.

The ggplot2 Package 000	Scatterplots 0000●0	Line Graphs 00000	Histograms 0000000	Boxplots 0000	Bar Charts 000
Overplotting III					

- Alternatively, can manipulate data set by jittering points a small random amount so that they no longer lie on top of each other.
- Consider the data set consisting of (0,0), (0,0), (0,0), (0,0) and (1,1):

The ggp1ot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	000000	00000	0000000	0000	000
Overplotting III					

- Alternatively, can manipulate data set by jittering points a small random amount so that they no longer lie on top of each other.
- Consider the data set consisting of (0,0), (0,0), (0,0), (0,0) and (1,1):

```
ggplot(data = jiggle_df, mapping = aes(x = x, y = y)) +
geom_point()
```



The ggp1ot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	000000	00000	0000000	0000	000
Overplotting III					

- Alternatively, can manipulate data set by jittering points a small random amount so that they no longer lie on top of each other.
- Consider the data set consisting of (0,0), (0,0), (0,0), (0,0) and (1,1):

```
ggplot(data = jiggle_df, mapping = aes(x = x, y = y)) +
geom_point()
```



It looks like there are just 2 observations!

The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	00000●	00000	0000000	0000	000
Overplotting III					

- Alternatively, can manipulate data set by jittering points a small random amount so that they no longer lie on top of each other.
- Consider the data set consisting of (0,0), (0,0), (0,0), (0,0) and (1,1):

```
ggplot(data = jiggle_df, mapping = aes(x = x, y = y)) +
geom_jitter(width = .05, height = .05)
```



The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	00000●	00000	0000000	0000	000
Overplotting III					

- Alternatively, can manipulate data set by jittering points a small random amount so that they no longer lie on top of each other.
- Consider the data set consisting of (0,0), (0,0), (0,0), (0,0) and (1,1):

```
ggplot(data = jiggle_df, mapping = aes(x = x, y = y)) +
geom_jitter(width = .05, height = .05)
```



 To jitter points, use the layer geom\_jitter(width = ..., height = ...) instead of geom\_points()

The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	000000	●0000	0000000	0000	000

## Section 3

Line Graphs

The ggplot2 Package 000	Scatterplots 000000	Line Graphs ○●○○○	Histograms 0000000	Boxplots 0000	Bar Charts 000
Changes over tim	e				

• How do bike use patterns change throughout the day?

The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	000000	O●000	0000000	0000	000
Changes over t	ime				

- How do bike use patterns change throughout the day?
- Consider the following summary information:

## # A tibble: 24 x 2 StartHour ## n ## <int> <int> 0 118 1 ## 2 ## 69 1 3 ## 2 50 ## 4 3 20 ## 5 4 35 6 5 71 ## 7 6 104 ## 8 7 270 ## 9 8 492 ## 392 ## 10 9 ## # ... with 14 more rows ## # i Use `print(n = ...)` to see more rows

The ggp1ot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	000000	00●00	0000000	0000	000
Line Graphs					

• Frequently, we compare two quantitative variables where one variable represents time. It is illustrative to connect neighboring points with a smooth curve.

The ggplot2 Package 000	Scatterplots 000000	Line Graphs 00●00	Histograms 0000000	Boxplots 0000	Bar Charts 000

### Line Graphs

- Frequently, we compare two quantitative variables where one variable represents time. It is illustrative to connect neighboring points with a smooth curve.
  - Compare the following:

The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	000000	○○●○○	0000000	0000	000

### Line Graphs

- Frequently, we compare two quantitative variables where one variable represents time. It is illustrative to connect neighboring points with a smooth curve.
  - Compare the following:



• These **line graphs** (or time series) provide stronger sequential and/or cyclic visual cues.

The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	000000	000●0	0000000	0000	000

### Making Line Graphs

To construct a line graph , use geom\_line() with the aesthetic mapping

```
aes(x = ..., y = ...).
ggplot(data = biketown2, mapping = aes(x = StartHour, y = n)) +
geom_line()
```



The ggplot2 Package 000	Scatterplots 000000	Line Graphs 0000●	Histograms 0000000	Boxplots 0000	Bar Charts 000
Combining Plots					

• We can also overlay points on the lines by *adding* a geom\_point layer!

```
ggplot(data = biketown2, mapping = aes(x = StartHour, y = n)) +
geom_line()+
geom_point()
```



The ggplot2 Package 000	Scatterplots 000000	Line Graphs 0000●	Histograms 0000000	Boxplots 0000	Bar Charts 000
Combining Plots					

• We can also overlay points on the lines by *adding* a geom\_point layer!

```
ggplot(data = biketown2, mapping = aes(x = StartHour, y = n)) +
geom_line()+
geom_point()
```



• Note that both geom\_line and geom\_point inherit the data and mapping arguments specified in the original ggplot function.

The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	000000	00000	●000000	0000	000

## Section 4

Histograms

The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	000000	00000	O●OOOOO	0000	000

#### The Distribution of a Variable

• Consider the Distance variable in the biketown data set. What are its minimum, maximum, and central values?

The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	000000	00000	O●00000	0000	000

#### The Distribution of a Variable

- Consider the Distance variable in the biketown data set. What are its minimum, maximum, and central values?
- What proportion of observations are "close" to these extremes?
| The ggplot2 Package | Scatterplots | Line Graphs | Histograms | Boxplots | Bar Charts |
|---------------------|--------------|-------------|------------|----------|------------|
| 000                 | 000000       | 00000       | O●00000    | 0000     | 000        |
|                     |              |             |            |          |            |

#### The Distribution of a Variable

- Consider the Distance variable in the biketown data set. What are its minimum, maximum, and central values?
- What proportion of observations are "close" to these extremes?
- These questions can be answered by exploring the distribution of a variable, which is
  a representation of the unique values it takes along with the frequency it takes them.

The ggplot2 Package 000	Scatterplots 000000	Line Graphs 00000	Histograms 00●0000	Boxplots 0000	Bar Charts 000
Histograms					

• Distributions are most commonly visualized by way of the histogram

The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	000000	00000	00●0000	0000	000

#### Histograms

- Distributions are most commonly visualized by way of the histogram
- To create a histogram:
  - Divide the x-axis into a sequence of equally-sized intervals (or bins).
  - For each, count the number of observations falling in that interval.
  - Draw bars with height equal to count and with width spanning the interval.

The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	000000	00000	00●0000	0000	000

#### Histograms

- Distributions are most commonly visualized by way of the histogram
- To create a histogram:
  - Divide the x-axis into a sequence of equally-sized intervals (or bins).
  - For each, count the number of observations falling in that interval.
  - Draw bars with height equal to count and with width spanning the interval.

```
ggplot(data = biketown_short, mapping = aes(x = Distance_Miles)) +
geom_histogram(bins = 50, color = "White")
```



The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	000000	00000	00●0000	0000	000

#### Histograms

- Distributions are most commonly visualized by way of the histogram
- To create a histogram:
  - Divide the x-axis into a sequence of equally-sized intervals (or bins).
  - For each, count the number of observations falling in that interval.
  - Draw bars with height equal to count and with width spanning the interval.

```
ggplot(data = biketown_short, mapping = aes(x = Distance_Miles)) +
geom_histogram(bins = 50, color = "White")
```



• Minimimum? Maximum? Center? Spread?

The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	000000	00000	000●000	0000	000

## The Shape of You (Distributions)

• Histograms also reveal qualitative information about the shape of a variable's distribution:

The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	000000	00000	000●000	0000	000

## The Shape of You (Distributions)

Histograms also reveal qualitative information about the shape of a variable's distribution:



The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	000000	00000	000●000	0000	000

## The Shape of You (Distributions)

Histograms also reveal qualitative information about the shape of a variable's distribution:



The ggplot2 Package 000	Scatterplots 000000	Line Graphs 00000	Histograms 0000●00	Boxplots 0000	Bar Charts 000
How many hing	:7				

• The number of bins used can radically affect the shape of the histogram.

The ggplot2 Package 000	Scatterplots 000000	Line Graphs 00000	Histograms 0000●00	Boxplots 0000	Bar Charts 000
How many bins?					

- The number of bins used can radically affect the shape of the histogram.
  - Use bins= to set the number of bins in a histogram

The ggplot2 Package 000	Scatterplots 000000	Line Graphs 00000	Histograms 0000●00	Boxplots 0000	Bar Charts 000
How many bin	s?				
• The number	of bins used car	n radically affect	the shape of th	e histogram.	

• Use bins= to set the number of bins in a histogram

```
ggplot(data = biketown_short, mapping = aes(x = Distance_Miles))+
geom_histogram(bins=10, color = "white")
```



The ggp1ot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	000000	00000	00000●0	0000	000
The Effect of B	in Size				

• Each of the following is a histogram for *the same data*, with different values for the bins = argument in geom\_histogram()

The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	000000	00000	00000●0	0000	000

### The Effect of Bin Size

• Each of the following is a histogram for *the same data*, with different values for the bins = argument in geom\_histogram()



The ggplot2 Package 000	Scatterplots 000000	Line Graphs 00000	Histograms 000000●	Boxplots 0000	Bar Charts 000
How many bins?					

• Alternatively, we can specify the width of bins using binwidth =

The ggp1ot2 Package 000	Scatterplots 000000	Line Graphs 00000	Histograms 000000●	Boxplots 0000	Bar Charts 000
How many bins	5?				
<ul> <li>Alternatively,</li> </ul>	, we can specify	the width of bir	ns using binwid	th =	
ggplot( <mark>data =</mark> bi geom_histogram	<pre>ketown_short, n (binwidth = 1,</pre>	<pre>mapping = aes(x color = "white</pre>	= Distance_Mi ")	les))+	
	4000 -				
	3000 -				
	2000 - 1000 -		_		
	0-				

Distance\_Miles

ggplot(data = biketown\_short, mapping = aes(x = Distance\_Miles))+
geom\_histogram(binwidth = 0.5, color = "white")

ò



The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	000000	00000	0000000	●000	000

# Section 5

Boxplots

The ggplot2 Package 000	Scatterplots 000000	Line Graphs 00000	Histograms 0000000	Boxplots O●OO	Bar Charts 000
Summary Statistic	CS				

• The **five-number summary** of a data set consists of: Minimum, 1st Quartile (Q1), Median, 3rd Quartile (Q3), Maximum.

The ggplot2 Package 000	Scatterplots 000000	Line Graphs 00000	Histograms 0000000	Boxplots O●OO	Bar Charts 000
Summary Statist	tics				

- The **five-number summary** of a data set consists of: Minimum, 1st Quartile (Q1), Median, 3rd Quartile (Q3), Maximum.
- The median is a value so that 50% of data lies above it and 50% lies below.

The ggplot2 Package 000	Scatterplots 000000	Line Graphs 00000	Histograms 0000000	Boxplots O●OO	Bar Charts 000
Summary Statisti	cs				

- The **five-number summary** of a data set consists of: Minimum, 1st Quartile (Q1), Median, 3rd Quartile (Q3), Maximum.
- The median is a value so that 50% of data lies above it and 50% lies below.
- The 1st / 3rd quartiles are values so that 25% / 75% of data lies below it and 75% / 25% lies above.

The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	000000	00000	0000000	0●00	000
Summary Statisti	CS				

- The **five-number summary** of a data set consists of: Minimum, 1st Quartile (Q1), Median, 3rd Quartile (Q3), Maximum.
- The median is a value so that 50% of data lies above it and 50% lies below.
- The 1st / 3rd quartiles are values so that 25% / 75% of data lies below it and 75% / 25% lies above.
- The median separates the data into two equal parts. Note Q1 is also the median of the lower part, while Q3 is the median of the upper part.

The ggp1ot2 Package 000	Scatterplots 000000	Line Graphs 00000	Histograms 0000000	Boxplots O●OO	Bar Charts 000
Summary Statisti	cs				

- The **five-number summary** of a data set consists of: Minimum, 1st Quartile (Q1), Median, 3rd Quartile (Q3), Maximum.
- The median is a value so that 50% of data lies above it and 50% lies below.
- The 1st / 3rd quartiles are values so that 25% / 75% of data lies below it and 75% / 25% lies above.
- The median separates the data into two equal parts. Note Q1 is also the median of the lower part, while Q3 is the median of the upper part.
- The interquartile range (IQR) is Q3 Q1 and measures the spread of the middle 50% of the data.

The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
OOO	000000	00000	0000000	O●OO	000
Summary Statisti	cs				

- The **five-number summary** of a data set consists of: Minimum, 1st Quartile (Q1), Median, 3rd Quartile (Q3), Maximum.
- The median is a value so that 50% of data lies above it and 50% lies below.
- The 1st / 3rd quartiles are values so that 25% / 75% of data lies below it and 75% / 25% lies above.
- The median separates the data into two equal parts. Note Q1 is also the median of the lower part, while Q3 is the median of the upper part.
- The interquartile range (IQR) is Q3 Q1 and measures the spread of the middle 50% of the data.
- Taken together, the five-number summary provides a measure of center and spread of a data set.

The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	000000	00000	0000000	00●0	000
Boxplots					

• The five-number summary can be visualized by way of the boxplot.

The ggplot2 Package 000	Scatterplots 000000	Line Graphs 00000	Histograms 0000000	Boxplots 00●0	Bar Charts 000
Boxplots					

- The five-number summary can be visualized by way of the boxplot.
- Consider the five number summary for Distance\_miles in the biketown data

The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	000000	00000	0000000	OO●O	000
Boxplots					

- The five-number summary can be visualized by way of the boxplot.
- Consider the five number summary for Distance\_miles in the biketown data

Min	Q1	Median	Q3	Max
0	0.79	1.48	2.68	23.75

The ggplot2 Package 000	Scatterplots 000000	Line Graphs 00000	Histograms 0000000	Boxplots 0000	Bar Charts 000

## Boxplots

- The five-number summary can be visualized by way of the boxplot.
- Consider the five number summary for Distance\_miles in the biketown data

Min	Q1	Median	Q3	Max
0	0.79	1.48	2.68	23.75



The ggplot2 Package 000	Scatterplots 000000	Line Graphs 00000	Histograms 0000000	Boxplots 00●0	Bar Charts 000

#### **Boxplots**

- The five-number summary can be visualized by way of the boxplot.
- Consider the five number summary for Distance\_miles in the biketown data

Min	Q1	Median	Q3	Max
0	0.79	1.48	2.68	23.75



- The Top / Bottom of box correspond to Q3 / Q1, while center line is median.
- The "whiskers" extend 1.5 × IQR in either direction from box edge.
- Outliers are any observations outside this range, and are plotted as points.

The ggplot2 Package 000	Scatterplots 000000	Line Graphs 00000	Histograms 0000000	Boxplots 000●	Bar Charts 000
Side-by-side Boxr	olots				
, ,					

• Often, we compare the distribution of a variable conditioned on values of a 2nd.

The ggplot2 Package 000	Scatterplots 000000	Line Graphs 00000	Histograms 0000000	Boxplots 000●	Bar Charts 000
Side-by-side Boxplots					

- Often, we compare the distribution of a variable conditioned on values of a 2nd.
- To do so, include an x-position aesthetic mapping from the 2nd variable.

The ggp1ot2 Package 000	Scatterplots 000000	Line Graphs 00000	Histograms 0000000	Boxplots 000●	Bar Charts 000
Side-by-side Boxplots					

- Often, we compare the distribution of a variable conditioned on values of a 2nd.
- To do so, include an x-position aesthetic mapping from the 2nd variable.
- To have boxes span horizontally, rather than vertically, add a coord\_flip() layer.

The ggp1ot2 Package 000	Scatterplots 000000	Line Graphs 00000	Histograms 0000000	Boxplots 000●	Bar Charts 000
Side-by-side Boxplots					

- Often, we compare the distribution of a variable conditioned on values of a 2nd.
- To do so, include an x-position aesthetic mapping from the 2nd variable.
- To have boxes span horizontally, rather than vertically, add a coord\_flip() layer.

ggplot(data = biketown, mapping = aes(x = PaymentPlan, y = Distance\_Miles)) +
geom\_boxplot()+ coord\_flip()



The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	000000	00000	0000000	0000	●00

## Section 6

Bar Charts

The ggplot2 Package 000	Scatterplots 000000	Line Graphs 00000	Histograms 0000000	Boxplots 0000	Bar Charts O●O
Bar Charts					

• Both Boxplots and Histograms show the distribution of *quantitative* variables.

The ggplot2 Package 000	Scatterplots 000000	Line Graphs 00000	Histograms 0000000	Boxplots 0000	Bar Charts O●O
Bar Charts					

- Both Boxplots and Histograms show the distribution of *quantitative* variables.
- We use Bar Charts to visualize the distribution of *categorical* variables, whose values are broken down into distinct levels.

The ggplot2 Package 000	Scatterplots 000000	Line Graphs 00000	Histograms 0000000	Boxplots 0000	Bar Charts O●O
Bar Charts					

- Both Boxplots and Histograms show the distribution of *quantitative* variables.
- We use Bar Charts to visualize the distribution of *categorical* variables, whose values are broken down into distinct levels.
- Investigate the distribution of bike use by month

The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	000000	00000	0000000	0000	O●O
Bar Charts					

- Both Boxplots and Histograms show the distribution of *quantitative* variables.
- We use Bar Charts to visualize the distribution of *categorical* variables, whose values are broken down into distinct levels.
- Investigate the distribution of bike use by month

```
ggplot(data = biketown, mapping = aes(x = Month)) +
geom_bar()
```


The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	000000	00000	0000000	0000	00●
Segmented ,	/ Stacked Bar Cł	narts			

• Bar charts used to visualize the *joint distribution* of 2 categorical variables.

The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	000000	00000	0000000	0000	00●

• Bar charts used to visualize the *joint distribution* of 2 categorical variables.



The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	000000	00000	0000000	0000	00●

• Bar charts used to visualize the *joint distribution* of 2 categorical variables.



 Each bar divided into count by fill variable.

The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	000000	00000	0000000	0000	00●

• Bar charts used to visualize the *joint distribution* of 2 categorical variables.



- Each bar divided into count by fill variable.
- Hard to make direct comparisons

The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	000000	00000	0000000	0000	00●

Bar charts used to visualize the joint distribution of 2 categorical variables.



- Each bar divided into count by fill variable.
- Hard to make direct comparisons

The ggplot2 Package	Scatterplots	Line Graphs	Histograms	Boxplots	Bar Charts
000	000000	00000	0000000	0000	00●

Bar charts used to visualize the joint distribution of 2 categorical variables.



- Each bar divided into count by fill variable.
- Hard to make direct comparisons

• Each bar divided into proportion by fill variable.