

Module 2: Displaying Data

Jon Starkweather, PhD

`jonathan.starkweather@unt.edu`

Consultant

Research and **Statistical Support**



Introduction to Statistics for the Social Sciences



The RSS short courses

The Research and Statistical Support (RSS) office at the University of North Texas hosts a number of “Short Courses”. A list of them is available at:

<http://www.unt.edu/rss/Instructional.htm>

Outline

- 1 Collecting Data
 - Context of an example study
 - Variable(s) of Interest
 - Demographic Variables

Outline

- 1 Collecting Data
 - Context of an example study
 - Variable(s) of Interest
 - Demographic Variables
- 2 Displaying Data in Tables
 - Frequency Tables
 - Stem and Leaf plots

Outline

- 1 Collecting Data
 - Context of an example study
 - Variable(s) of Interest
 - Demographic Variables
- 2 Displaying Data in Tables
 - Frequency Tables
 - Stem and Leaf plots
- 3 Figures and/or Graphs
 - Graphing Data
 - Bar Graphs
 - Histograms
 - Boxplots
 - Scatterplots

Outline

- 1 Collecting Data
 - Context of an example study
 - Variable(s) of Interest
 - Demographic Variables
- 2 Displaying Data in Tables
 - Frequency Tables
 - Stem and Leaf plots
- 3 Figures and/or Graphs
 - Graphing Data
 - Bar Graphs
 - Histograms
 - Boxplots
 - Scatterplots
- 4 Summary of Module 2
 - Additional Considerations/Issues
 - Summary of Module 2
 - What's next

Example Study

Do students in a statistics class *really* pay attention to the slides being presented?

Example Study

Do students in a statistics class *really* pay attention to the slides being presented?

We set up hidden digital video cameras to clearly record the faces (& facial expressions) of each student in class.

Example Study

Do students in a statistics class *really* pay attention to the slides being presented?

We set up hidden digital video cameras to clearly record the faces (& facial expressions) of each student in class.

Next, we alter a typical slide presentation, so that in the middle of the presentation (middle of class time) a single, high-resolution image appears for 5 seconds, then disappears.

Example Study

Do students in a statistics class *really* pay attention to the slides being presented?

We set up hidden digital video cameras to clearly record the faces (& facial expressions) of each student in class.

Next, we alter a typical slide presentation, so that in the middle of the presentation (middle of class time) a single, high-resolution image appears for 5 seconds, then disappears.

- The image chosen displays a 90-year-old man wearing only bright red lipstick and a red string-bikini.

Example Study

Do students in a statistics class *really* pay attention to the slides being presented?

We set up hidden digital video cameras to clearly record the faces (& facial expressions) of each student in class.

Next, we alter a typical slide presentation, so that in the middle of the presentation (middle of class time) a single, high-resolution image appears for 5 seconds, then disappears.

- The image chosen displays a 90-year-old man wearing only bright red lipstick and a red string-bikini.

Using the videos, the time (100ths of a second) between the image appearing and any visible reaction from each of the students was recorded.

Data Collected: Reaction Time

Table 1: Raw Data

| Students 1-10 | Students 11-20 | Students 21-30 | Students 31-40 |
|---------------|----------------|----------------|----------------|
| 60 | 55 | 55 | 57 |
| 50 | 53 | 51 | 56 |
| 62 | 57 | 56 | 52 |
| 61 | 59 | 56 | 58 |
| 59 | 54 | 54 | 53 |
| 59 | 58 | 55 | 56 |
| 57 | 56 | 54 | 51 |
| 57 | 60 | 56 | 55 |
| 58 | 58 | 53 | 56 |
| 52 | 55 | 57 | 56 |

Reaction Time in 100ths of a second.

Demographic Variables

Generally, when we collect data, we are also interested in the individuals we collect it from; be those individuals: persons, schools, corporations, etc.

Demographic Variables

Generally, when we collect data, we are also interested in the individuals we collect it from; be those individuals: persons, schools, corporations, etc.

So, we will often collect information on demographic variables which help us understand the nature of the individuals (i.e., demographic variables tell us something about our sample).

Demographic Variables

Generally, when we collect data, we are also interested in the individuals we collect it from; be those individuals: persons, schools, corporations, etc.

So, we will often collect information on demographic variables which help us understand the nature of the individuals (i.e., demographic variables tell us something about our sample).

- With human participants: age, gender/sex, ethnicity, location, income or socio-economic status (SES)...etc.

Demographic Variables

Generally, when we collect data, we are also interested in the individuals we collect it from; be those individuals: persons, schools, corporations, etc.

So, we will often collect information on demographic variables which help us understand the nature of the individuals (i.e., demographic variables tell us something about our sample).

- With human participants: age, gender/sex, ethnicity, location, income or socio-economic status (SES)...etc.

In our current example, we might also take note of the students' gender and their class standing.

Demographic Variables

Generally, when we collect data, we are also interested in the individuals we collect it from; be those individuals: persons, schools, corporations, etc.

So, we will often collect information on demographic variables which help us understand the nature of the individuals (i.e., demographic variables tell us something about our sample).

- With human participants: age, gender/sex, ethnicity, location, income or socio-economic status (SES)...etc.

In our current example, we might also take note of the students' gender and their class standing.

- Male (0) or Female (1)

Demographic Variables

Generally, when we collect data, we are also interested in the individuals we collect it from; be those individuals: persons, schools, corporations, etc.

So, we will often collect information on demographic variables which help us understand the nature of the individuals (i.e., demographic variables tell us something about our sample).

- With human participants: age, gender/sex, ethnicity, location, income or socio-economic status (SES)...etc.

In our current example, we might also take note of the students' gender and their class standing.

- Male (0) or Female (1)
- Freshman (1), Sophomore (2), Junior (3), Senior (4)

Data: What do we have?

Our data consists of 40 individuals or cases.

Data: What do we have?

Our data consists of 40 individuals or cases.

- In this example; humans, specifically students in a statistics class

Data: What do we have?

Our data consists of 40 individuals or cases.

- In this example; humans, specifically students in a statistics class

Our data also consists of three variables.

Data: What do we have?

Our data consists of 40 individuals or cases.

- In this example; humans, specifically students in a statistics class

Our data also consists of three variables.

- Reaction: the reaction time of each of the students expressed in 100ths of a second.

Data: What do we have?

Our data consists of 40 individuals or cases.

- In this example; humans, specifically students in a statistics class

Our data also consists of three variables.

- Reaction: the reaction time of each of the students expressed in 100ths of a second.
- Gender: male or female

Data: What do we have?

Our data consists of 40 individuals or cases.

- In this example; humans, specifically students in a statistics class

Our data also consists of three variables.

- Reaction: the reaction time of each of the students expressed in 100ths of a second.
- Gender: male or female
- Class Standing: Freshman, Sophomore, Junior, Senior

Data: What do we have?

Our data consists of 40 individuals or cases.

- In this example; humans, specifically students in a statistics class

Our data also consists of three variables.

- Reaction: the reaction time of each of the students expressed in 100ths of a second.
- Gender: male or female
- Class Standing: Freshman, Sophomore, Junior, Senior

So what do we do with our data now?

Displaying the data

So, what do we do with our data now?

Displaying the data

So, what do we do with our data now?

- Generally, the *raw* data is not terribly useful.

Displaying the data

So, what do we do with our data now?

- Generally, the *raw* data is not terribly useful.
- So, the first thing we do is **display** the data so that we can begin to interpret it.

Displaying the data

So, what do we do with our data now?

- Generally, the *raw* data is not terribly useful.
- So, the first thing we do is **display** the data so that we can begin to interpret it.
- There are many ways of displaying data

Displaying the data

So, what do we do with our data now?

- Generally, the *raw* data is not terribly useful.
- So, the first thing we do is **display** the data so that we can begin to interpret it.
- There are many ways of displaying data
 - We usually start with frequency **tables**

Displaying the data

So, what do we do with our data now?

- Generally, the *raw* data is not terribly useful.
- So, the first thing we do is **display** the data so that we can begin to interpret it.
- There are many ways of displaying data
 - We usually start with frequency **tables**
 - Then use the frequencies to create some **figures**.

Tables & Figures

A key distinction to make when displaying data (and talking about displaying data) is the distinction between Tables & Figures.

Tables & Figures

A key distinction to make when displaying data (and talking about displaying data) is the distinction between Tables & Figures.

- A **Table** is made up of straight lines and text content which can be typeset.

Tables & Figures

A key distinction to make when displaying data (and talking about displaying data) is the distinction between Tables & Figures.

- A **Table** is made up of straight lines and text content which can be typeset.
 - letters and numbers

Tables & Figures

A key distinction to make when displaying data (and talking about displaying data) is the distinction between Tables & Figures.

- A **Table** is made up of straight lines and text content which can be typeset.
 - letters and numbers
 - Table 1 above, which showed our raw data

Tables & Figures

A key distinction to make when displaying data (and talking about displaying data) is the distinction between Tables & Figures.

- A **Table** is made up of straight lines and text content which can be typeset.
 - letters and numbers
 - Table 1 above, which showed our raw data
- A **Figure** is any graphical representation

Tables & Figures

A key distinction to make when displaying data (and talking about displaying data) is the distinction between Tables & Figures.

- A **Table** is made up of straight lines and text content which can be typeset.
 - letters and numbers
 - Table 1 above, which showed our raw data
- A **Figure** is any graphical representation
 - Wiring schematic, photograph, thermal image, MRI image, topographical map, ...

Tables & Figures

A key distinction to make when displaying data (and talking about displaying data) is the distinction between Tables & Figures.

- A **Table** is made up of straight lines and text content which can be typeset.
 - letters and numbers
 - Table 1 above, which showed our raw data
- A **Figure** is any graphical representation
 - Wiring schematic, photograph, thermal image, MRI image, topographical map, ...
 - And common graphs as we will see in a moment.

Frequency Tables

- There are many kinds of frequency tables & figures, each is used to display how *frequently* (how many times) each score occurred, commonly called the “distribution” of scores.

Frequency Tables

- There are many kinds of frequency tables & figures, each is used to display how *frequently* (how many times) each score occurred, commonly called the “distribution” of scores.
- Frequency **tables** display:

Frequency Tables

- There are many kinds of frequency tables & figures, each is used to display how *frequently* (how many times) each score occurred, commonly called the “distribution” of scores.
- Frequency **tables** display:
 - All the possible values of a variable sequentially in the first column

Frequency Tables

- There are many kinds of frequency tables & figures, each is used to display how *frequently* (how many times) each score occurred, commonly called the “distribution” of scores.
- Frequency **tables** display:
 - All the possible values of a variable sequentially in the first column
 - Then they display the *frequency of occurrence* of each value in the next column.

Frequency Tables

- There are many kinds of frequency tables & figures, each is used to display how *frequently* (how many times) each score occurred, commonly called the “distribution” of scores.
- Frequency **tables** display:
 - All the possible values of a variable sequentially in the first column
 - Then they display the *frequency of occurrence* of each value in the next column.
 - Frequencies are calculated by simply counting the number of individuals having the specified value.

Frequency Tables

- There are many kinds of frequency tables & figures, each is used to display how *frequently* (how many times) each score occurred, commonly called the “distribution” of scores.
- Frequency **tables** display:
 - All the possible values of a variable sequentially in the first column
 - Then they display the *frequency of occurrence* of each value in the next column.
 - Frequencies are calculated by simply counting the number of individuals having the specified value.
- Occasionally, the cumulative frequency and/or cumulative percentage is listed in subsequent columns.

Frequency Table: Continuous Variable

Table 2: Reaction Time Frequency Table

| Reaction Time | Frequency |
|---------------|-----------|
| 50 | 1 |
| 51 | 2 |
| 52 | 2 |
| 53 | 3 |
| 54 | 3 |
| 55 | 5 |
| 56 | 8 |
| 57 | 5 |
| 58 | 4 |
| 59 | 3 |
| 60 | 2 |
| 61 | 1 |
| 62 | 1 |

Frequency Table: Categorical Variable

Table 3: Class Standing Frequency Table

| Class Standing | Code | Frequency | Percent | Cum.Percent |
|----------------|------|-----------|---------|-------------|
| Freshman | 1 | 5 | 12.5 | 12.5 |
| Sophomore | 2 | 12 | 30.0 | 42.5 |
| Junior | 3 | 16 | 40.0 | 82.5 |
| Senior | 4 | 7 | 17.5 | 100 |

Optional percentages and cumulative percentages displayed.

Frequency Table: Categorical Variable

Table 4: Gender Frequency Table

| Gender | Code | Frequency |
|--------|------|-----------|
| Male | 0 | 14 |
| Female | 1 | 26 |

Stem and Leaf Plots

Stem and Leaf plots can be thought of as *grouped* frequency tables. Each 'stem' is the part of an individual's score which is associated with a group, and each 'leaf' is the part associated with the individual.

Table 5: Reaction Time Stem and Leaf plot

| Frequency | Stem | Leaf |
|-----------|------|---------------------------|
| 3 | 5 | 0 1 1 |
| 5 | 5 | 2 2 3 3 3 |
| 8 | 5 | 4 4 4 5 5 5 5 5 |
| 13 | 5 | 6 6 6 6 6 6 6 6 7 7 7 7 7 |
| 7 | 5 | 8 8 8 9 9 9 9 |
| 3 | 6 | 0 0 1 |
| 1 | 6 | 2 |

Stem and Leaf Plots

A more representative example of a Stem and Leaf plot. Each stem represents a decade.

Table 6: Age Stem and Leaf plot

| Frequency | Stem | Leaf |
|-----------|------|-----------------------|
| 21 | 2 | 011222333344555667899 |
| 16 | 3 | 1112233455677889 |
| 14 | 4 | 00122345556789 |
| 13 | 5 | 2233455556667 |
| 5 | 6 | 00126 |
| 3 | 7 | 126 |
| 1 | 8 | 2 |

Graphing Data

- Displaying data in Frequency tables is fairly easy and informative when:

Graphing Data

- Displaying data in Frequency tables is fairly easy and informative when:
 - There are relatively few case (individuals)

Graphing Data

- Displaying data in Frequency tables is fairly easy and informative when:
 - There are relatively few case (individuals)
 - There are relatively few possible values (of the variable).

Graphing Data

- Displaying data in Frequency tables is fairly easy and informative when:
 - There are relatively few case (individuals)
 - There are relatively few possible values (of the variable).
- Often the values of a variable are much more or much less numerous, resulting in uninformative frequency tables.

Graphing Data

- Displaying data in Frequency tables is fairly easy and informative when:
 - There are relatively few case (individuals)
 - There are relatively few possible values (of the variable).
- Often the values of a variable are much more or much less numerous, resulting in uninformative frequency tables.
- It is almost always beneficial to go beyond the frequency tables and create some sort of graphical display to make the data more interpretable.

Bar Graphs

Bar Graphs are used to graphically display frequencies and are used when the variable is categorical (discrete).

Bar Graphs

Bar Graphs are used to graphically display frequencies and are used when the variable is categorical (discrete). Recall our demographic variables from our example.

Bar Graphs

Bar Graphs are used to graphically display frequencies and are used when the variable is categorical (discrete). Recall our demographic variables from our example.

- Gender

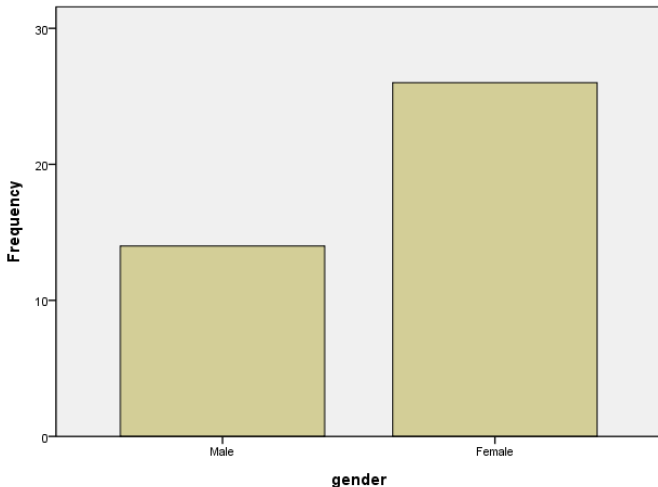
Bar Graphs

Bar Graphs are used to graphically display frequencies and are used when the variable is categorical (discrete). Recall our demographic variables from our example.

- Gender
- Class Standing

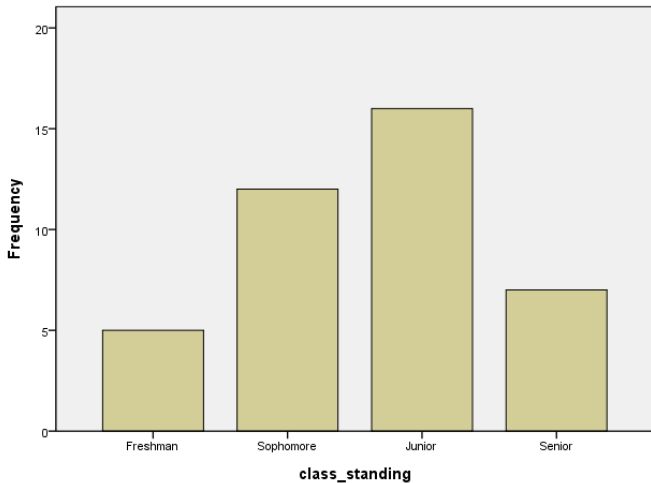
Bar Graph of Gender

Figure 1: Gender Bar Graph



Bar Graph of Class Standing

Figure 2: Class Standing Bar Graph



Histograms

Histograms are used to graphically display frequencies and are especially useful when the variable's values are numerous.

Histograms

Histograms are used to graphically display frequencies and are especially useful when the variable's values are numerous. Recall, our example measured reaction time in 100ths of a second.

Histograms

Histograms are used to graphically display frequencies and are especially useful when the variable's values are numerous. Recall, our example measured reaction time in 100ths of a second.

- If we had over a thousand students in our example *and*

Histograms

Histograms are used to graphically display frequencies and are especially useful when the variable's values are numerous. Recall, our example measured reaction time in 100ths of a second.

- If we had over a thousand students in our example *and*
- We had students distributed between a very late reaction of 450 (4.5 seconds) and a very quick reaction of 3 (.03 of a second).

Histograms

Histograms are used to graphically display frequencies and are especially useful when the variable's values are numerous. Recall, our example measured reaction time in 100ths of a second.

- If we had over a thousand students in our example *and*
- We had students distributed between a very late reaction of 450 (4.5 seconds) and a very quick reaction of 3 (.03 of a second).

Then it would be beneficial to use a Histogram.

Histograms

Histograms are used to graphically display frequencies and are especially useful when the variable's values are numerous. Recall, our example measured reaction time in 100ths of a second.

- If we had over a thousand students in our example *and*
- We had students distributed between a very late reaction of 450 (4.5 seconds) and a very quick reaction of 3 (.03 of a second).

Then it would be beneficial to use a Histogram.

- Looks like a bar graph; but the 'bars' have no space between them, *and*

Histograms

Histograms are used to graphically display frequencies and are especially useful when the variable's values are numerous. Recall, our example measured reaction time in 100ths of a second.

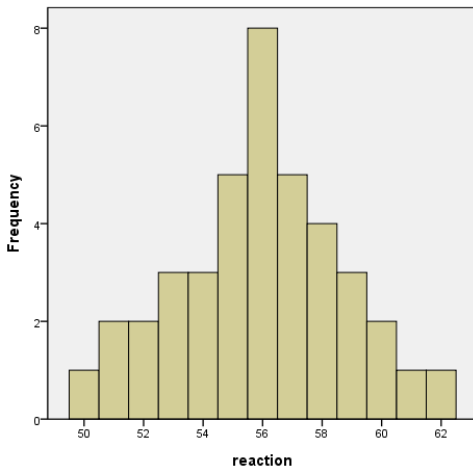
- If we had over a thousand students in our example *and*
- We had students distributed between a very late reaction of 450 (4.5 seconds) and a very quick reaction of 3 (.03 of a second).

Then it would be beneficial to use a Histogram.

- Looks like a bar graph; but the 'bars' have no space between them, *and*
- Each bar represents multiple values of the variable.

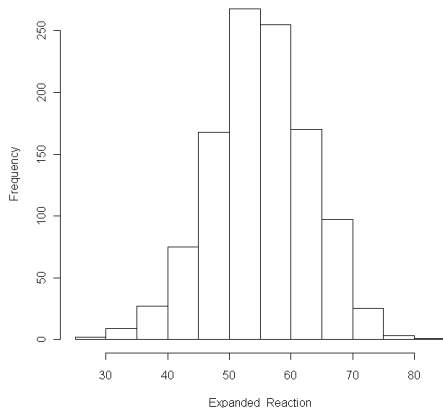
Histogram of Reaction Time

Figure 3: Reaction Time Histogram w/example study data



Histogram of a much large and more widely distributed sample of Reaction Time

Figure 4: Expanded Histogram (not example study data)



Pros and Cons of Histograms

- **Pro 1:** Makes a large distribution of scores easy to interpret graphically.

Pros and Cons of Histograms

- **Pro 1:** Makes a large distribution of scores easy to interpret graphically.
- **Pro 2:** Makes spotting *outliers* easy.

Pros and Cons of Histograms

- **Pro 1:** Makes a large distribution of scores easy to interpret graphically.
- **Pro 2:** Makes spotting *outliers* easy.
 - **Outlier:** an extreme case, a case whose value is far to one end or another of the distribution.

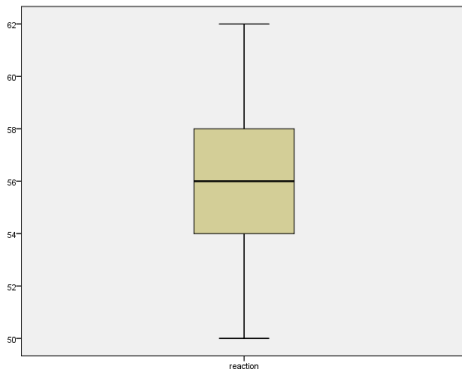
Pros and Cons of Histograms

- **Pro 1:** Makes a large distribution of scores easy to interpret graphically.
- **Pro 2:** Makes spotting *outliers* easy.
 - **Outlier:** an extreme case, a case whose value is far to one end or another of the distribution.
- **Con 1:** Some information is lost when bars represent more than one value.

Boxplots

Boxplots are good for showing the where the bulk of data lies in relation to the tails (whiskers) of a distribution.

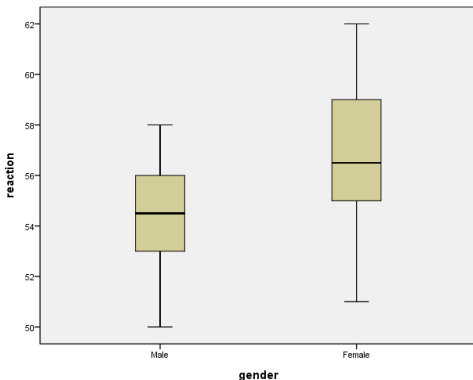
Figure 5: Reaction Time Boxplot



Multiple Group boxplots

Here, the boxplot shows the reaction time distribution of both males and females.

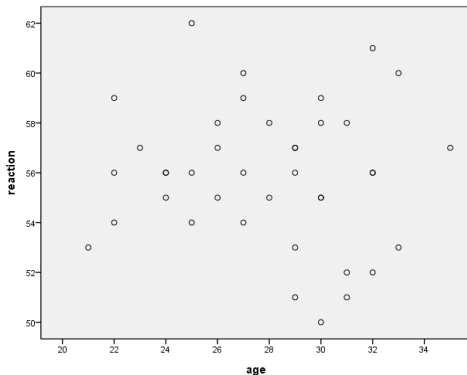
Figure 6: Reaction Time by Gender Boxplot



Scatterplots

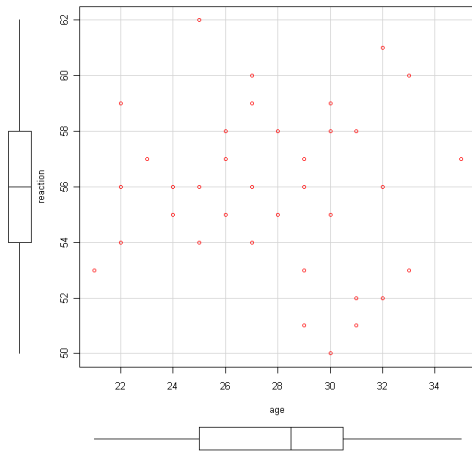
Scatterplots are used to show how two (or more) variables are distributed together. Here, the 'plain' scatterplot shows reaction time and age.

Figure 7: Reaction Time and Age



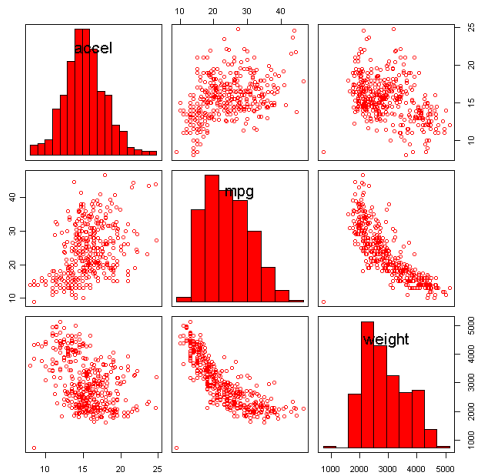
Scatterplots

Figure 8: Reaction Time and Age w/boxplots



Scatterplot Matrix: more than 2 variables

Figure 9: Scatterplot Matrix w/histograms on diagonal



Additional Thoughts on Displaying Data

- Using tables and figures makes the data more interpretable, or accessible, for ourselves and others.

Additional Thoughts on Displaying Data

- Using tables and figures makes the data more interpretable, or accessible, for ourselves and others.
- While producing graphs is incredibly easy with computers, we must not forget that understanding the data and conveying it are the goals of displaying data.

Additional Thoughts on Displaying Data

- Using tables and figures makes the data more interpretable, or accessible, for ourselves and others.
- While producing graphs is incredibly easy with computers, we must not forget that understanding the data and conveying it are the goals of displaying data.
 - **Not simply producing nice graphs**

Additional Thoughts on Displaying Data

- Using tables and figures makes the data more interpretable, or accessible, for ourselves and others.
- While producing graphs is incredibly easy with computers, we must not forget that understanding the data and conveying it are the goals of displaying data.
 - **Not simply producing nice graphs**
- When creating graphs; be very careful about how you scale each axis.

Additional Thoughts on Displaying Data

- Using tables and figures makes the data more interpretable, or accessible, for ourselves and others.
- While producing graphs is incredibly easy with computers, we must not forget that understanding the data and conveying it are the goals of displaying data.
 - **Not simply producing nice graphs**
- When creating graphs; be very careful about how you scale each axis.
 - There is great power when creating a simple graph.

Additional Thoughts on Displaying Data

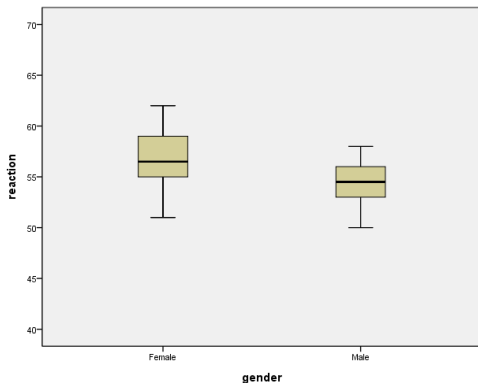
- Using tables and figures makes the data more interpretable, or accessible, for ourselves and others.
- While producing graphs is incredibly easy with computers, we must not forget that understanding the data and conveying it are the goals of displaying data.
 - **Not simply producing nice graphs**
- When creating graphs; be very careful about how you scale each axis.
 - There is great power when creating a simple graph.
- An example follows.

Scale Importance

Here is the same exact information as displayed in Figure 6.

- Changing the scale has made the two genders *look* more similar by compressing the boxplots.

Figure 10: Reaction Time by Gender Boxplot



Summary of Module 2

Module 2 covered the following topics:

- Collecting Data

Summary of Module 2

Module 2 covered the following topics:

- Collecting Data
 - Context of an example study

Summary of Module 2

Module 2 covered the following topics:

- Collecting Data
 - Context of an example study
 - Variable(s) of interest

Summary of Module 2

Module 2 covered the following topics:

- Collecting Data
 - Context of an example study
 - Variable(s) of interest
 - Demographic variables.

Summary of Module 2

Module 2 covered the following topics:

- Collecting Data
 - Context of an example study
 - Variable(s) of interest
 - Demographic variables.
- Displaying Data in Tables.

Summary of Module 2

Module 2 covered the following topics:

- Collecting Data
 - Context of an example study
 - Variable(s) of interest
 - Demographic variables.
- Displaying Data in Tables.
 - Frequency Tables

Summary of Module 2

Module 2 covered the following topics:

- Collecting Data
 - Context of an example study
 - Variable(s) of interest
 - Demographic variables.
- Displaying Data in Tables.
 - Frequency Tables
 - Stem & Leaf plots

Summary of Module 2

Module 2 covered the following topics:

- Collecting Data
 - Context of an example study
 - Variable(s) of interest
 - Demographic variables.
- Displaying Data in Tables.
 - Frequency Tables
 - Stem & Leaf plots
- Displaying Data in Graphs

Summary of Module 2

Module 2 covered the following topics:

- Collecting Data
 - Context of an example study
 - Variable(s) of interest
 - Demographic variables.
- Displaying Data in Tables.
 - Frequency Tables
 - Stem & Leaf plots
- Displaying Data in Graphs
 - Bar Graphs

Summary of Module 2

Module 2 covered the following topics:

- Collecting Data
 - Context of an example study
 - Variable(s) of interest
 - Demographic variables.
- Displaying Data in Tables.
 - Frequency Tables
 - Stem & Leaf plots
- Displaying Data in Graphs
 - Bar Graphs
 - Histograms

Summary of Module 2

Module 2 covered the following topics:

- Collecting Data
 - Context of an example study
 - Variable(s) of interest
 - Demographic variables.
- Displaying Data in Tables.
 - Frequency Tables
 - Stem & Leaf plots
- Displaying Data in Graphs
 - Bar Graphs
 - Histograms
 - Boxplots

Summary of Module 2

Module 2 covered the following topics:

- Collecting Data
 - Context of an example study
 - Variable(s) of interest
 - Demographic variables.
- Displaying Data in Tables.
 - Frequency Tables
 - Stem & Leaf plots
- Displaying Data in Graphs
 - Bar Graphs
 - Histograms
 - Boxplots
 - Scatterplots

Summary of Module 2

Module 2 covered the following topics:

- Collecting Data
 - Context of an example study
 - Variable(s) of interest
 - Demographic variables.
- Displaying Data in Tables.
 - Frequency Tables
 - Stem & Leaf plots
- Displaying Data in Graphs
 - Bar Graphs
 - Histograms
 - Boxplots
 - Scatterplots
- Additional Considerations & Issues

This concludes Module 2

Next time Module 3.

- Next time we'll begin covering descriptive statistics.
- Until next time; have a nice day.

These slides initially created on: September 14, 2010

These slides last updated on: September 23, 2010

- The bottom date shown is the date this Adobe.pdf file was created; \LaTeX ¹ has a command for automatically inserting the date of a document's creation.

¹This document was created in \LaTeX using the Beamer package