BIOM5010:

Statistics #2B

Reference:

onlinestatbook.com/
Chapter 2A, 2B2, 2B4
Graphics can be hard to understand


(C) 2011 – 2017. Andy Adler
Advantage Little Guys

Big banks have widely slashed the availability of free checking since 2009

- Banks > $50B
- All Other Banks
- Credit Unions

<table>
<thead>
<tr>
<th>Year</th>
<th>Banks &gt; $50B</th>
<th>All Other Banks</th>
<th>Credit Unions</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>96.0%</td>
<td>78.0%</td>
<td>85.0%</td>
</tr>
<tr>
<td>2011</td>
<td>70.8%</td>
<td>34.6%</td>
<td>78.6%</td>
</tr>
</tbody>
</table>

Source: Moebes.Services

Proportion offering Free Checking

- Credit Unions
- Other Banks
- Banks > $50 B

2009 2011
**Example: Consider these data**

<table>
<thead>
<tr>
<th>Computer</th>
<th>#</th>
<th>Relative. Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>85</td>
<td>0.17</td>
</tr>
<tr>
<td>Windows</td>
<td>60</td>
<td>0.12</td>
</tr>
<tr>
<td>Mac</td>
<td>355</td>
<td>0.71</td>
</tr>
<tr>
<td>Total</td>
<td>500</td>
<td>1.00</td>
</tr>
</tbody>
</table>
- avoid 3D effects unless necessary
Problem Graphs #2

Lie Factor = \frac{\text{Size of Effect in Graphic}}{\text{Size of Real Effect}}

- increased area makes effect appear larger
Problem Graphs #3

Lie Factor = \(\frac{\text{Size of Effect in Graphic}}{\text{Size of Real Effect}}\)

- starting axis above zero makes effect appear larger
Pie Charts

Source: Fox News (http://www.datavis.ca/gallery/images/fox-news-piechart.jpg)
Pie and Bar charts

- When can you use?
- When not appropriate?

- must add to 100%
- don't show in 3D
- multiple pie charts confusing
- can't show error bars

- can show error bars
- be careful of problems in next pages
Comparing Distributions

- This is OK, it doesn't mislead
- However, what is the goal? what is the message?
- Do we want to show that more people play on Wed than Sun?
- Do we want to show which games are preferred?
- lines connecting poker - blackjack etc makes it look like the variables are connected. But they are not.
[from junkcharts] message-first. the designer should figure out what message to convey, then design the visualization to convey that message.
Message-first
Box Plots

- onlinestatbook.com/2/graphing_distributions/boxplots.html

- Median
- 50th percentile
- 25th percentile
- 75th percentile
- 1.5xH
- 1.5xH
- Data point just below 1.5xH
- Data point just above 1.5xH
- Outlier(s)
Box Plots

- onlinestatbook.com/2/graphing_distributions/boxplots.html
Questions

• (Final 2015, Q1) Recovery times were compared for patients undergoing gastric bypass surgery using traditional techniques vs. patients undergoing the same procedure using the new DrRobot robotic surgery platform. This study examined the recovery times for two groups of patients at two hospitals in Los Angeles, California: A) 25 patients who received robotic surgery at the (private) University Hospital, and B) 25 patients who received traditional surgery at the (public) Sisters of Mercy Hospital.
  - (1d) Discuss one source of possible bias in this study. How might it affect the results?
(Final 2015, Q2) Based on the results of the first study (in Q#1), your colleagues in marketing create an image to illustrate the mean recovery time (in days) of traditional techniques vs. the new DrRobot robotic surgery system. Do you feel comfortable that this is a fair representation? Discuss three issues.