Traditional Charting

- Manual Methods can teach us a lot.
- Charting and Graphing tell a story visually...
- ...but all charts have inherent bias.
- Sometimes the bias is accidental...
- ...but sometimes the bias is intentional.
- Watch out!

Has anyone ever seen this chart?

EXAMPLE #1
OLD STYLE CHARTING

Napoleon’s March

- This is considered by many the best statistical chart ever drawn.
- It was drawn in 1869 by French civil engineer Charles Minard.
- It shows Napoleon’s march on Moscow in the European War of 1812, and the retreat.
- It shows six independent channels of information with a minimum of “chart junk”.

#1: Strength of Army

Width of Bands = Number of Men.
Shows flanking maneuvers, places where groups rejoin.
EXAMPLE #2
MODERN CHARTING

#2 & #3: Location of Army
Latitude, Longitude over time

#4 & #5: Direction of Army
$(\Delta X, \Delta Y)$ over time

#6: Temperature (“Reammur”)
Bottom temperature scale tied to retreat. Damn cold! Many men froze on the march.

This is a Great Chart!
- It shows in poignant detail what happened to the French army. 422,000 men left Poland, only 10,000 returned.
- But as good as it is, it doesn’t answer one very important question:
- What was the Russian army doing all this time?
- While it is an honest chart, it is by its nature biased.

U.S. Presidential Election 2004
This is what we saw on election night (Blue=John Kerry, Red=George W. Bush):

http://www.personal.umich.edu/~mejn/election/2004/
Images available under Creative Commons license
OK. What’s Wrong with it?

• Is this truth?
• Is this all of the truth?
• Or does it show just one version of the truth?

Here’s the original map again

Pretty understandable, isn’t it?

Scale by Population

Wyoming nearly vanishes, Massachusetts becomes significantly larger.

Scale by Electoral College Votes

Similar to Previous Map.

What about on the county level?

Counties now red or blue. But Malheur and Harney counties in Eastern Oregon are each the size of Massachusetts, with combined population less than Amherst plus UMass!

Scale County Size by Population

Malheur and Harney counties vanish, even though their land areas are very large!
But No County is All Red or All Blue

Color-Scale counties between red or blue dependent on proportion of votes for each candidate.

Then Scale by Population

We are not red or blue, but purple!

What does this show us?

• Every version of the map is “true”.
• Some versions show information in a more complete manner, but...
• ...some versions are easier to understand than others.
• Thought experiment: can you imagine the public reaction if the last map was shown on TV election night?

EXAMPLE #3
INTENTIONAL BIAS

Intentionally biased chart #1

What’s Wrong?

• Scale on right-hand side does not start at 0%; it goes from 34% to 42%, not from 0% to 42%!
• Hides parts of the bars between 0% and 34%.
• Inflates the difference between the heights, making the discrepancy appear much bigger than it really is.
• Can be unintentional, but look for it! It might be someone trying to mislead you.
What’s Wrong?

- The scale is upside down!
- Visual bias fools us into thinking the “stand your ground” law causes a drop in gun deaths.
BUT WHAT ABOUT EXCEL?

The Take-Away

• We can’t do the same tricks in Excel just shown, but...
• …we can observe many of the same design principles:
  – Tell a story simply,
  – Add nothing unnecessary,
  – Tell the whole story, but...
  – …be aware the story will always be incomplete,
  – Tell the truth.

Here’s a Sample Spreadsheet

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Q1</td>
<td>Q2</td>
<td>Q3</td>
<td>Q4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Fred</td>
<td>15000</td>
<td>17000</td>
<td>16000</td>
<td>17000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Sam</td>
<td>8000</td>
<td>6000</td>
<td>7000</td>
<td>8000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Mary</td>
<td>25000</td>
<td>21000</td>
<td>23000</td>
<td>22000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Column Chart

• Everyone else calls this a Bar Chart
• Three series (Fred, Sam, Mary)
• Groups related items from different series together (compare all Q1s, all Q2s, etc., see how Fred did relative to Sam relative to Mary)

Annotated Column Chart
Annotated Column Chart

- Most labels are optional, but:
  - You need a title or you don’t know the story you are telling,
  - You need to know the scales on X and Y,
  - You need to know what the X and Y axes mean,
  - You need a legend or the data are meaningless (unless there is only one series)

Column Chart (Data Rotated)

- Same data, just different interpretation
- Four series now (Q1, Q2, Q3, Q4)
- Makes sense to see all Fred’s data together, all Sam’s data together, and all Mary’s data together.
- Rotating a column chart usually makes sense.

Line Chart

- Same data, just different interpretation
- “Connects the dots”
- Shows trends over time.
**Line Chart (Rotated)**
- Rotating the data changes the interpretation again, but...
- Is Fred to Sam to Mary a time sequence?
- No, this makes no sense!

**Stacked Column Chart**
- All of Fred’s data are in one stack, all of Sam’s in another, all of Mary’s in a third.
- You get a visual total for free.
- Easy to tell at a glance that Sam is doing the worst of all three people.

**3D Column Chart**
- Looks very pretty.
- Does it convey additional information?
- No, it actually makes things worse:
  - Hard to tell actual height of bars
  - Perspective effects distort interpretation
  - Can hide short values in back

**3D Pie Chart**
3D Pie Chart

• Looks very pretty.
• Does it convey additional information?
• No, it actually makes things worse:
  – Hard to tell size of pie wedges
  – Perspective foreshortens wedges on sides
  – Perspective widens wedges on front (and back)

Conclusion

• 3D is pretty.
• Don’t use it!
• 2D almost always conveys your story better.