Presenting Data

Chapter 2
Outline

• Types of Charts
• Game Analytics Examples
• Guidelines for Charts
Tables

- Generally, **independent** variable in left column and **dependent** variables next

### Table 1. Number of student on campus and off by year

<table>
<thead>
<tr>
<th>Class rank</th>
<th>Do you live on campus?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Off-campus</td>
</tr>
<tr>
<td>Freshman</td>
<td>37</td>
</tr>
<tr>
<td>Sophomore</td>
<td>42</td>
</tr>
<tr>
<td>Junior</td>
<td>90</td>
</tr>
<tr>
<td>Senior</td>
<td>62</td>
</tr>
<tr>
<td>Total</td>
<td>231</td>
</tr>
</tbody>
</table>

**Checklist**
- Number and caption
- Units labeled (as appropriate)
- Minimal vertical lines (or none)
- Lines only to break apart areas (or use Bold)

Make sure to consider *message*. Often much clearer in chart!
“Right” Chart Depends on **Variable Type**

- **Qualitative** (Categorical) variables
  - Can have states or subclasses
    - e.g., position: [striker, goalie, midfield]
  - Can be ordered or unordered
    - e.g., bronze, silver, gold → **ordered**
    - e.g., support, warrior, specialist → **unordered**

- **Quantitative** (Numeric) variables
  - Numeric levels
  - **Discrete** or continuous
    - e.g., goals in season, speed in meters
    - e.g., takedowns, win percentage
Categorical: Bar Chart (1 of 2)

- Chart containing rectangles ("bars") where length represents count, amount, or percent (aka "column chart")
- Better than table for comparing numbers

Olympic Games Hosted

<table>
<thead>
<tr>
<th>Continent</th>
<th>Olympic Games Hosted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Europe</td>
<td>15</td>
</tr>
<tr>
<td>North America</td>
<td>5</td>
</tr>
<tr>
<td>Asia</td>
<td>2</td>
</tr>
<tr>
<td>Australia</td>
<td>1</td>
</tr>
<tr>
<td>South America</td>
<td>1</td>
</tr>
</tbody>
</table>

https://www.kristakingmath.com/blog/bar-graphs-and-pie-charts
Categorical: Bar Chart (2 of 2)

- Chart containing rectangles ("bars") where length represents count, amount, or percent

[Diagram of bar chart showing categories Europe, North America, Asia, Australia, South America with corresponding values for Summer games and Winter games.]
Categorical: Pareto Chart

• Bar chart, arranged most to least frequent
• Line showing cumulative percent
• Helps identify most common

Sort by column D (Data -> Sort high to low)
New column E for percent [=D2/SUM(D$2:D$11)]
Note: $ “locks” value in (e.g., D$2 versus D2)
New column F for running [=SUM(E$2:E2)]
Select B, D and F. Insert “combo chart”
Categorical: Pie Chart

- Wedge-shaped areas ("pie slices") - represent count, amount or percent of each category from whole
- Compare relative amounts at a glance
- Best if few slices since quantifying "size" of pie difficult
- Comparing pies also difficult

"The Effects of Latency and Jitter on a First Person Shooter: Team Fortress 2"
http://www.cs.wpi.edu/~claypool/iqp/tf2/
Histogram

Bar chart for grouped numerical data
- No gaps between adjacent bars

Select GPA data
Insert → Statistics Chart → Histogram
Can adjust bins, overflow/underflow

# kills winners make in PUBG

https://www.mathsisfun.com/data/images/bar-chart-vs-histogram.gif
https://theparttimeanalyst.com/2018/12/10/pubg-eda/
Time Series Plot

• Associate data with date
• Line graph with dates (proportionally spaced!)

http://www.soundandvision.com/content/violence-and-video-games

Two Variables – Scatter Plot

- Two numerical variables, one on each axis
- Reveal patterns in relationship
- Setup “right” models (later)

“Intelligent Simulation of Worldwide Application Distribution for OnLive’s Server Network”

http://www.cs.wpi.edu/~claypool/mqp/onlive/

Select two of {win, pick, ban}
Insert ➔ scatter plot
Radar Chart (1 of 2)

- Also called “star”, “spider”, “web” or “kiviat” charts
- Good for quick visual comparison, especially when axes unequal

Select top line {win, pick, ban} + 3 rows (Ctrl-select)
Insert ➔ Other ➔ Radar scatter plot

Note: need to normalize data to scale Axes
Insert column E (“B Norm”)
=E2/MAX(E$2:E42)
Copy and paste down
Radar Charts (2 of 2)

- Note, axes themselves hard to read values
- Value is to compare shapes

https://www.data-to-viz.com/caveat/spider.html
Many More Charts!

https://en.wikipedia.org/wiki/Chart

- Bubble
- Waterfall
- Tree
- Gap
- Polar
- Violin
- Candlestick
- Kagi
- Gantt
- Nolan
- Pert
- Smith
- Skyline
- Vowel
- Nomogram
- Natal

- If common chart effective for message, use
- Otherwise, learn/use other charts as needed
- But remember – may need to explain how to read
Outline

• Types of Charts (done)
• Game Analytics Examples (next)
• Guidelines for Charts
Game Analytics Visual Analysis


- Player choices (e.g., build units)
- Density of activities (e.g., where spend time on map)
- Movement through levels
Figure 1. Pie-charts show which types of towers have been built on the different building lots. The radius of the pie-chart is proportional to the number of towers built (Kayali, et al., 2014).

(Custom game, comparative study)
Figure 2. (a) Heatmap of death locations on the Team Fortress 2 map Goldrush. (b) Heatmap showing locations where players of a tower defense game collected coins dropped by defeated enemies (Kayali, et al., 2014).
Player Location – Heat Map (2 of 2)

Assassin’s Creed

Where play testers failed

Result: Make red areas easier

Note, Heat Map for Tables, Too!

Excel tutorial at: [https://trumpexcel.com/heat-map-excel/](https://trumpexcel.com/heat-map-excel/)

<table>
<thead>
<tr>
<th></th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2  January</td>
<td>600</td>
<td>708</td>
<td>594</td>
</tr>
<tr>
<td>3  February</td>
<td>607</td>
<td>984</td>
<td>749</td>
</tr>
<tr>
<td>4  March</td>
<td>901</td>
<td>886</td>
<td>908</td>
</tr>
<tr>
<td>5  April</td>
<td>608</td>
<td>615</td>
<td>835</td>
</tr>
<tr>
<td>6  May</td>
<td>715</td>
<td>833</td>
<td>734</td>
</tr>
<tr>
<td>7  June</td>
<td>520</td>
<td>663</td>
<td>618</td>
</tr>
<tr>
<td>8  July</td>
<td>731</td>
<td>521</td>
<td>950</td>
</tr>
<tr>
<td>9  August</td>
<td>709</td>
<td>663</td>
<td>987</td>
</tr>
<tr>
<td>10 September</td>
<td>633</td>
<td>863</td>
<td>979</td>
</tr>
<tr>
<td>11 October</td>
<td>533</td>
<td>651</td>
<td>841</td>
</tr>
<tr>
<td>12 November</td>
<td>996</td>
<td>958</td>
<td>749</td>
</tr>
<tr>
<td>13 December</td>
<td>792</td>
<td>717</td>
<td>875</td>
</tr>
</tbody>
</table>

Red means sales are low.
(game: *Infinite Mario*, clone of *Super Mario Bros.*)

**Figure 4.** Examples of path visualizations coupled with color-coding to communicate additional information. Top: color coding reflects the reported expertise of players obtained through a pre-game survey. Middle: colors depict the state in which the player’s character currently resides in. Bottom: the color-gradient reflects physiological data measured in the form of galvanic skin response (Mirza-Babaei, et al., 2014).
Figure 5. Left: Player movement between regions, cities, and battlegrounds on the World of Warcraft continent Outland. Right: Corresponding matrix view with cells colored according to the number of players moving from one area to another.
Outline

• Types of Charts  (done)
• Game Analytics Examples  (done)
• Guidelines for Charts  (next)

https://xkcd.com/1945/
Guidelines for Good Charts (1 of 7)

• Require minimum effort from reader
  – Perhaps *most* important metric
  – Given two, can pick one that takes less reader effort

![Diagram](image-url)
Guidelines for Good Charts (2 of 7)

• Maximize information
  – Make self-sufficient
  – Key words in place of symbols
    • e.g., “Gold IV” and not “Player A”
    • e.g., “Daily Games Played” not “Games Played”
  – Axis labels as informative as possible
    • e.g., “Game Time (seconds)” not “Game Time”
  – Help by using captions (or title, if stand-alone)
    • e.g., “Game time in seconds versus player skill in total hours played”

http://www.phplot.com/phplotdocs/conc-labels.html
Guidelines for Good Charts (3 of 7)

• Minimize ink

https://www.slideshare.net/NicoleMarinsek/darkhorse-line-chart
Guidelines for Good Charts (4 of 7)

• Use commonly accepted practices
  – Present what people expect
  – e.g., origin at (0,0)
  – e.g., independent (cause) on x-axis, dependent (effect) on y-axis
  – e.g., x-axis scale is linear
  – e.g., increase left to right, bottom to top
  – e.g., scale divisions equal, proportional

• Departures are permitted but require extra effort from reader → so use sparingly!
Guidelines for Good Charts (5 of 7)

• Avoid ambiguity
  – Show coordinate axes
    • at right angles
  – Show origin
    • usually at (0,0)
  – Identify individual curves and bars
    • With key/legend or label
  – Do not plot multiple variables on same chart
    • Single y-axis
Guidelines for Good Charts (6 of 7)

• Don’t connect categorical data with lines
  – Lines joining successive points signify that they can be approximately interpolated
  – If don’t have meaning, should not use line chart

- No linear relationship between champion types
- Instead, use column chart
- Don’t connect with lines
Guidelines for Good Charts (6 of 7)

- Can deceive as easily as can convey meaning

**EXAMPLE 2:** Amount of Land Planted with Grapes for the Wine Industry.

- Missing x-axis (1997 too far over)
- Missing y-axis hard to compare (1950 height > 1970)
- Data points unclear
Groupwork
Improving Bad Charts

2. Work as a Team: Compare lists. Discuss.
3. Write down combined.

Checklist for Good Charts

• Axes
  – Are both axes labeled?
  – Are the axis labels self-explanatory and concise?
  – Are the scale and divisions shown on both axes?
  – Are the min and max ranges appropriate?
  – Are the units indicated?

• Lines/Curves/Points
  – Is the number of lines/curves reasonably small?
  – Are curves labeled?
  – Are all symbols clearly distinguishable?
  – Is a concise, clear legend provided?
  – Does the legend obscure any data?

• Information
  – If the y-axis is variable, is an indication of spread (error bars) shown?
  – Are grid lines required to read data (if not, then remove)?

• Scale
  – Are units increasing left to right (x-axis) and bottom to top (y-axis)?
  – Do all charts use the same scale?
  – Are the scales contiguous?
  – Is bar chart order systematic?
  – Are bars appropriate width, spacing?

• Overall
  – Does the whole chart add information to reader?
  – Are there no curves/symbols/text that can be removed and still have the same information?
  – Does the chart have a title or caption (not both)?
  – Is the chart self-explanatory and concise?
  – Do the variables plotted give more information than alternatives?
  – Is chart referenced and discussed in any accompanying report?
Describing Chart in Report & Presentation

• “Formula”
  – Describe all axes
    • E.g., “The x-axis is time since game began, in seconds”
  – Describe data sets/trendlines
    • E.g., “The blue dots are the average maze completion time”
  – **Then** provide message
    • E.g., “Notice how the red bar is higher than the blue, indicating that ...”
• Example on Web page

[Image of chart with data analysis]

http://web.cs.wpi.edu/~imgd2905/d23/samples/analysis-example.html
Guidelines for Good Charts (Summary)

• For each chart, go over “checklist”
• The more “yes” answers, the better
  – Remember, while guidelines, art and not science
  – So, may consciously decide not to follow these guidelines if better without them → but have good reason!
• In practice, takes several trials before arriving at “best” chart
• Want to present message the most: accurately, simply, concisely, logically
• Accompany with description! Text or verbal
  – Remember, audience/reader has not seen! – Make sure to introduce