Types of Variables

- Qualitative (Categorical) variables
  - Can have states or subclasses
    - e.g., rank: [platinum, diamond, gold]
  - Can be ordered or unordered
    - e.g., bronze, silver, gold ➔ ordered
    - e.g., support, tank, jungler ➔ unordered
- Quantitative (Numeric) variables
  - Numeric levels
    - Discrete or continuous
    - e.g., gold per minute, deaths, character level
    - e.g., kills + assists / deaths ratio, win percentage

Outline

- Types of Charts
- Guidelines for Charts
- Common Mistakes

Categorical: Bar Chart

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

Categorical: Pareto Chart

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

Categorical: Pie Chart

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

Note: bars could be sideways, too

Demo: imgdpops.xlsx
Categorical: Cross-Classification Table

- Multi-column table that presents count or percent for 2+ categorical variables
  - Good for comparison across multi-categorical data

<table>
<thead>
<tr>
<th>Class rank</th>
<th>Freshman</th>
<th>Sophomore</th>
<th>Junior</th>
<th>Senior</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>37</td>
<td>183</td>
<td>137</td>
<td></td>
<td>357</td>
</tr>
<tr>
<td></td>
<td>42</td>
<td>48</td>
<td>90</td>
<td></td>
<td>180</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>89</td>
<td>90</td>
<td></td>
<td>279</td>
</tr>
<tr>
<td></td>
<td>62</td>
<td>1</td>
<td>63</td>
<td></td>
<td>138</td>
</tr>
</tbody>
</table>

- Demo: grades.xlsx

Numeric: Frequency Distribution

- Groups of numeric values and frequency
  - e.g., Survey of Champion “skins” bought with RP
  - 1, 2, 1, 0, 3, 4, 0, 1, 1, 2, 2, 3, 2, 1, 4, 0, 0
  - Cluster into groups
  - Report frequency per group

<table>
<thead>
<tr>
<th>Skins</th>
<th>Freq</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4</td>
<td>20%</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>30%</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>25%</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>15%</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>10%</td>
</tr>
</tbody>
</table>

- May include percentage
  - Typically equal size
  - Sometimes ends are open (for extremes)
  - Bin size/number variable
  - Too many and not readable
  - Guide:
    - 100 or less: 7-10
    - 101-200: 11-15
    - 200+: 13-20

Cumulative Distribution

- Cumulative amount of data with value or less
  - Easy to see min, max, median
  - Compare shapes of distributions

- Demo: lol-patches.xlsx

Histogram

- Bar chart for grouped numerical data
  - No (or small) gaps between adjacent bars

- Ages of professional League players

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

Stem and Leaf Display

- “Histogram-lite” for analysis w/out software
  - e.g., exam scores: 34, 81, 75, 51, 82, 96, 55, 66, 95, 87, 82, 88, 99, 50, 85, 72

| 9 | 6 5 9 |
| 8 | 1 2 7 2 8 5 |
| 7 | 5 2 |
| 6 | 6 |
| 5 | 1 5 0 |
| 4 | |
| 3 | 4 |

Time Series Plot

- Associate data with date
- Line graph with dates (proportionally spaced)

- Demo: majors.xlsx

Select year and majors
- Insert → Line Chart
- More Line Charts
Scatter Plot

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

![Scatter Plot Example](image)

“Intelligent Simulation of Worldwide Application Distribution for OnLive’s Server Network”
http://www.cs.wpi.edu/~claypool/mqp/onlive/

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

Radar Plot

- Also called “star charts” or “kiviat plots”
- Good for quick visual compare, especially when axes unequal

![Radar Plot Example](image)

Many More Charts!

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

- If common chart effective for message, use
- Learn/use other charts as needed

Game Analytics Charts


http://dl.acm.org/citation.cfm?id=2812792

- Player choices (e.g., build units)
- Density of activities (e.g., where spend time on map)
- Movement through levels

Player Locations – Heat Map (1 of 2)

![Player Location Heat Map Example](image)

Player Choices – Pie-Chart

![Player Choices Pie-Chart Example](image)
**Outline**

- Types of Charts (done)
- Guidelines for Charts (next)
  - Again, “art” not “rules”. Learn with experience. Recognize good/bad when see it.
- Common Mistakes

**Guidelines for Good Charts (1 of 5)**

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


Assassin's Creed
Where play testers failed
Result: Make red areas easier

![Movement (1 of 2)](game: Infinite Mario, clone of Super Mario Bros.)

Class: Infinite Mario
Build road to veterinary house
Shows exploration, where stuck

![Player Behavior - Node-link](Game: DOGeometry - build road to veterinary house)

Shows exploration, where stuck

![Guidelines for Good Charts (1 of 5)](require minimum effort from reader)

- Perhaps most important metric
  - Given two, can pick one that takes less reader effort
Guidelines for Good Charts (2 of 5)

- 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”
  - Axes 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”

Guidelines for Good Charts (3 of 5)

- Minimize ink (1 of 2)
  - Maximize information-to-ink ratio
  - Too much unnecessary ink makes chart cluttered, hard to read
    - e.g., no gridlines unless needed to help read
  - Chart that gives easier-to-read for same data is preferred

- Minimize ink (2 of 2)

Guidelines for Good Charts (4 of 5)

- 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
  - Departures are permitted, but require extra effort from reader → so use sparingly!

Guidelines for Good Charts (5 of 5)

- 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

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/legend 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

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

Outline

- Types of Charts (done)
- Guidelines for Charts (done)
- Common Mistakes (next)

Common Mistakes (1 of 6)

- Presenting too many alternatives on one chart
- Guidelines
  - More than 5 to 7 messages is too many
    - (Maybe related to the limit of human short-term memory?)
  - Line chart with 6+ curves
  - Column chart with 10+ bars
  - Pie chart with 8+ components
  - Each cell in histogram fewer than 5 values

Common Mistakes (2 of 6)

- Presenting many y-variables on single chart
  - Better to make separate graphs
  - Plotting many y-variables saves space, but better to require reader to figure out relationship
  - Sometimes, space constraints (e.g., journal/conference papers),
    - So may “bend” but better to remove than “break”

Common Mistakes (3 of 6)

- Using symbols in place of text
  - More difficult to read symbols than text
  - Reader must flip through report to see symbol mapping to text
  - Even if “save” writers time, really “wastes” it since reader is likely to skip!
Common Mistakes (4 of 6)

- Placing extraneous information on chart
  - Goal to convey message, so extra information distracting
  - e.g., Using gridlines only when exact values needed
  - e.g., Showing “per-user” data when only average user data needed

Common Mistakes (5 of 6)

- Selecting scale ranges improperly
  - Most prepared by automatic rules
  - Give good first-guess
  - But
    - May include outlying data points, shrinking body
    - May have endpoints hard to read since on axis
    - May place too many (or too few) tics
    - In practice, almost always over-ride scale values

Common Mistakes (6 of 6)

- Using line chart instead of column chart
  - Lines joining successive points signify that they can be approximately interpolated
  - If don’t have meaning, should not use line chart

Misleading Charts

Non-Zero Origins to Emphasize

- Normally, both axes meet at origin
- By moving and scaling, can magnify (or reduce) difference

Non-Zero Origins to Emphasize

- MINE
- YOURS

Which graph is better?
Non-Zero Origins to Emphasize (3 of 3)

- Choose scale so that vertical height of highest point is at least ¾ of the horizontal offset of right-most point
  - Three-quarters rule
- (And represent origin as 0,0)

Using Double-Whammy Graph

- Two curves can have twice as much impact
  - But if two metrics are related, knowing one predicts other ... so use one!

Plotting Quantities without Measure of Spread

- When random quantification, representing mean (or median) alone (or single data point!) not enough

Pictograms Scaled by Height

- If scaling pictograms, do by area not height since eye drawn to area
  - e.g., twice as good → doubling height quadruples area

Using Inappropriate Cell Size in Histogram

- Getting cell size “right” always takes more than one attempt
  - If too large, all points in same cell
  - If too small, lacks smoothness

Using Broken Scales in Column Charts

- By breaking scale in middle, can exaggerate differences
  - May be trivial, but then looks significant
  - Similar to “zero origin” problem
Pictorial Games (1 of 2)

- Can deceive as easily as can convey meaning

![Diagram 1: Australian Wine Exports to the United States](image1)

Pictorial Games (2 of 2)

- Can deceive as easily as can convey meaning

![Diagram 2: Amount of Land Planted with Grapes for the Wine Industry](image2)