Data Visualization

Making Charts & Graphs and Other Cool Stuff

CFDR Fall 2022 Workshop Series
Today’s Presentation

1. Why charts?
2. What are the different types of charts (with dos and don’ts)?
3. What are the basic principles of chart design?
4. What are some charts to be cautious of?
5. What are some programs (other than Excel) I can use to create data visualizations?
6. What are some outlets for data visualization publication?
7. Do you always need a chart?
Why — charts?
Designing good charts, however, presents more challenges than tabular display as it draws on the talents of both the scientist and the artist. You have to know and understand your data, but you also need a good sense of how the reader will visualize the chart’s graphical elements.”

~ Gary Klass
Picture Superiority Effect

Information is better remembered in tests of recall and item recognition when presented as pictures rather than words.
What are –

the different types of charts?
Histograms

A vertical bar chart that depicts the distribution of a set of data

**Characteristics**

- Bars represent the frequency of occurrence by classes of data.
- Enables you to see the shape of the data’s distribution.
- Like a bar chart, but a histogram groups numbers into ranges.
- The horizontal axis is continuous like a number line (no gaps between columns).
- Great way to show results of continuous data:
  - Weight
  - Height
  - How much time
- A Frequency Histogram uses vertical columns to show how many times each score occurs.

**Example**

```
Scores: 1,1,2,2,2,2,3,3,3,3,4,4,5
```

Source of chart: http://www.mathsisfun.com/data/histograms.html
Histograms

A vertical bar chart that depicts the distribution of a set of data

Clustered Column - Unformatted

<table>
<thead>
<tr>
<th>Scores</th>
<th>Frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

Histogram - Formatted

Scores: 1,1,2,2,2,2,3,3,3,3,4,4,5
Pie Charts

Generally used to show percentage or proportional data classified into nominal or ordinal categories

Characteristics

- Show the size of items in one data series, proportional to the sum of the items
- Useful for displaying data that are classified into nominal or ordinal categories
- Rules for pie charts:
  - Avoid using pie charts
  - Use only for data that add up to some meaningful total
  - Avoid comparisons across multiple pie charts
  - Five is the maximum number of slices, but two is better…

Example

Top Reasons for Fathers Leaving the Workforce in 2008

- School/Training, 20%
- Layoff, 19%
- Childcare, 15%
- Other, 46%
Prevalence of Pre-union First Birth across Demographic Characteristics

- Women: 11%
- Men: 9%
- Whites: 7%
- Hispanics: 13%
- Blacks: 33%
- B.A.: 2%
- Assoc. Deg.: 8%
- H.S.: 12%
- GED: 20%
- None: 19%

Prevalence of Pre-union First Birth by Race/Ethnicity:

- Blacks
  - Yes: 33%
  - No: 67%
- Hispanics
  - Yes: 7%
  - No: 93%
- Whites
  - Yes: 13%
  - No: 87%

Pie Charts

Generally used to show percentage or proportional data classified into nominal or ordinal categories

Pie Chart - Unformatted

Top Reasons for Fathers Leaving the Workforce in 2008

Pie Chart - Formatted

Top Reasons for Fathers Leaving the Workforce in 2008

- Other, 46%
- School/Training, 20%
- Layoff, 19%
- Childcare, 15%
Other Types of Pie Charts

**Doughnut**

Percent of young adults who enroll in a 4-year program by degree earned by age 25

- Bachelor’s degree: 49%
- Associate’s degree: 7%
- Did not finish: 44%


**Pie-of-Pie**

Percent of births by informal marital status of mother, 2005-2010

- Married: 57%
- Unmarried: 43%
- Cohabiting: 25%
- Single: 18%

Source: NSFG 2006-2010
Another Way to Use Pie Charts

• Not the primary form of distilling the information.
• Used as a graphic element to re-emphasize the focus of the Profile.
Column & Bar Charts

Useful for showing data changes over a period of time or for illustrating comparisons among items

Characteristics

- In column charts, categories are typically organized along the horizontal axis and values along the vertical axis.
- Allows user to compare values across categories. You can use a clustered column chart type when you have categories that represent:
  - Ranges of values (e.g., item counts)
  - Specific scale arrangements (e.g., a Likert scale)
  - Names that are not in any specific order (e.g., item names, geographic names, names of people)

Example – Side-by-Side Column Chart

Percentage of Same-Sex Couple Households with Minor Children by Sex of Couple and Race/Ethnicity of Household Head

Source: U.S. Census Bureau, American Community Survey, 1-Year Estimates, 2012
Column & Bar Charts

Useful for showing data changes over a period of time or for illustrating comparisons among items

Characteristics

- In column charts, categories are typically organized along the horizontal axis and values along the vertical axis.
- Allows user to compare values across categories. You can use a clustered column chart type when you have categories that represent:
  - Ranges of values (e.g., item counts)
  - Specific scale arrangements (e.g., a Likert scale)
  - Names that are not in any specific order (e.g., item names, geographic names, names of people)
- Stacked column: show the relationship of individual items to the whole

Example – Stacked Bars & Trend Lines

Changes in the Shares of Births to Single and Cohabitng Mothers Under Age 40

In column charts, categories are typically organized along the horizontal axis and values along the vertical axis.

- Allows user to compare values across categories. You can use a clustered column chart type when you have categories that represent:
  - Ranges of values (e.g., item counts)
  - Specific scale arrangements (e.g., a Likert scale)
  - Names that are not in any specific order (e.g., item names, geographic names, names of people)

- Stacked column: show the relationship of individual items to the whole
- 100% stacked column: compare the percentage that each value contributes to a total across categories.

Example – Stacked Bars & Trend Lines

Changes in the Shares of Births to Single, Cohabiting, and Married Mothers Under Age 40

Column & Bar Charts

Useful for showing data changes over a period of time or for illustrating comparisons among items

Simple Column Chart - Unformatted

Example – Simple Column Chart

Fathers Living with All of Their Children Race, Ethnicity & Nativity

- All fathers
- White
- Black
- NB Hispanic
- FB Hispanic

Fathers Living with All of Their Children Race, Ethnicity & Nativity

- All fathers: 74%
- White: 80%
- Black: 49%
- NB Hispanic: 62%
- FB Hispanic: 70%

Source: NSFG 2006-2010
Line Charts

Ideal for showing trends over time at equal intervals

Characteristics

- Time is almost always displayed on the X-axis from left to right.
- Spacing between markers in the x-axis should be proportional
- If you can’t easily see the pattern of each series, you may have too many
- Don’t use a legend – directly label the series, instead

Example – Simple Line Chart

Share of Married Mothers Experiencing a Premarital Birth by Race and Marriage Cohort

Source: The Integrated Fertility Survey Series (IFSS) is a project of the Population Studies Center and the Inter-university Consortium for Political and Social Research at the University of Michigan, with funding from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), grant 5R01 HD053533; Pamela J. Smock, PI.
Line Charts

Ideal for showing trends over time at equal intervals

Characteristics

- Time is almost always displayed on the X-axis from left to right.
- Don’t use a legend – directly label the series, instead.
- If you can’t easily see the pattern of each series, you may have too many.
- Spacing between markers in the x-axis should be proportional.
- Beware of scaling effects.
- When displaying fiscal or monetary data over-time, it is often best to use deflated data (e.g., inflation-adjusted or % of GDP).

Example – Line Chart with a Double Axis

Annual HMI Spending and Marriage & Divorce Rates, 2000 - 2010

Scatter Plots

Commonly used to show the relationship between two variables e.g., correlation

**Characteristics**
- Use two interval-level variables
- Fully define the variables with the axis titles
- If there is an implied causal relationship between the variables, place the independent variable on the X-axis and the dependent variable on the Y-axis
- Scale the axes to maximize the use of the plot area for displaying the data points
- It’s a good idea to add data labels to identify the cases
- In scatter plots, use empty circles as markers to let the reader see the overlapping points

**Example – Scatter Plot**

State Math Scores and Students’ TV Viewing Habits

Source: National Center for Educational Statistics, 1994
**Line Graph vs. Scatter Plots**

Can be used for trend data at UNEQUAL intervals

### Women’s Marriage to Divorce Ratio, 1970-2015

<table>
<thead>
<tr>
<th>Year</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1970</td>
<td>3.0</td>
</tr>
<tr>
<td>1980</td>
<td>2.0</td>
</tr>
<tr>
<td>1990</td>
<td>2.1</td>
</tr>
<tr>
<td>2000</td>
<td>2.4</td>
</tr>
<tr>
<td>2008</td>
<td>1.7</td>
</tr>
<tr>
<td>2009</td>
<td>1.8</td>
</tr>
<tr>
<td>2010</td>
<td>1.7</td>
</tr>
<tr>
<td>2011</td>
<td>1.7</td>
</tr>
<tr>
<td>2012</td>
<td>1.7</td>
</tr>
<tr>
<td>2013</td>
<td>1.8</td>
</tr>
<tr>
<td>2014</td>
<td>2.0</td>
</tr>
<tr>
<td>2015</td>
<td>2.1</td>
</tr>
</tbody>
</table>

Line Graph vs. Scatter Plots

Can be used for trend data at UNEQUAL intervals

Line with Markers

Scatter with Lines and Markers
Line Graph vs. Scatter Plots

Can be used for trend data at UNEQUAL intervals

Scatter with Lines and Markers - Unformatted

Scatter with Lines and Markers - Formatted

Women’s Marriage to Divorce Ratio, 1970-2015
Area Charts

Show percentage or proportional data classified into nominal or ordinal categories over time

Characteristics

- Use these in moderation. Fall victim to the same visualization problems as pie- and doughnut-charts…pies and doughnuts are also area charts!
  - We avoid them “whenever possible because visual perception in humans can only compare areas as rough estimates” (Few, 2012; p275).

- The primary issue with area charts is our brains tend to read them as line charts—disregarding the “area” aspect of the chart.
  - The example here isn’t as problematic as a regular ole area chart because it is a 100% stacked chart.

Example – Stacked Area Chart


Area Charts

Show percentage or proportional data classified into nominal or ordinal categories over time

Example – Area Chart

- Not stacked.
- Our brains will disregard the area underneath and behind.

Source: NCFMR analysis of Survey of Income and Program Participation, 2021
What are the – basic principles of chart design?
1. Simplify

- Sort data in a meaningful way
- Minimize ink-to-data ratio → remove unneeded chart elements
  - Gridlines
  - Chart borders
  - Axis titles
  - Legends
  - Markers and data labels
  - Decimal points (in axis and data labels)
  - Trend lines
  - NO 3D CHARTS
1. Simplify

NO 3D CHARTS!!!
2. Color vs. Black & White

ACCESSIBILITY is your #1 Concern!

- When in doubt → Black & white
- BUT color can help tell a story
  - Color = branding
    - Use a cohesive and consistent color palette
    - Be mindful of how your audience will view your chart(s)
      - Excel vs. Word vs. PDF
      - Color vs. B&W print copy
      - Colorblind audience
3. Do NOT Use Distorted Charts

- Do NOT misrepresent your data!
  - Use appropriate and consistent axis and scales
4. Present Related Charts Simultaneously

- One-after-another or side-by-side if possible
  - Use appropriate axis and scales

Figure 3. Remarriage Rates by Age Groups and Gender, 1990
Figure 4. Remarriage Rates by Age Groups and Gender, 2015

Source: Centers for Disease Control and Prevention, National Center for Health Statistics, Vital Stats and U.S. Census Bureau, American Community Survey, 2015
5. Know Your Audience

- Academics vs. lay people
- Undergraduate students vs. graduate students
- Graduate students vs. professors
- PAA presentation vs. job talk
6. TMC = TMI

- Too many charts (TMC) is as bad as too much information (TMI)...

  → Do NOT overload your audience!
What are some –

charts to be cautious of?
Charts to be Cautious of...

- Pie/ Doughnut Charts
- Area Charts
- Radar Charts

Example Radar chart by Visiblox
Charts to be Cautious of…

- Pie/Doughnut Charts
- Area Charts
- Radar Charts
- Circle Charts

Divorce Rates per 1,000 by Age Groups, 2015

Source: U.S. Census Bureau, American Community Survey, 2015 1-year est.
Charts to be Cautious of…

**Circle Chart**

**Divorce Rates per 1,000 by Age Groups, 2015**

**Column/ Bar Chart**

Vs.

Source: U.S. Census Bureau, American Community Survey, 2015 1-year est.
Charts to be Cautious of...

- Pie/Doughnut Charts
- Area Charts
- Radar Charts
- Circle Charts
- Unit Charts
Charts to be Cautious of...

- Pie/ Doughnut Charts
- Area Charts
- Radar Charts
- Circle Charts
- Unit Charts
- Funnel Charts

Source: https://peltiertech.com/bad-graphics-funnel-chart/
Charts to be Cautious of…

✓ Pie/ Doughnut Charts
✓ Area Charts
✓ Radar Charts
✓ Circle Charts
✓ Unit Charts
✓ Funnel Charts

- Waterfall Charts

What are some programs –
(other than Excel)
I can use to create
data visualizations?
Other Programs for Creating Data Visualizations

- Piktochart - [https://piktochart.com/](https://piktochart.com/)
  [https://create.piktochart.com/output/a9c32242c8ba-falcon-flames-copy-2]
Other Programs for Creating Data Visualizations

- Piktochart - https://piktochart.com/
- Canva - https://www.canva.com/

Other Programs for Creating Data Visualizations

- Piktochart - https://piktochart.com/
- Canva - https://www.canva.com/
- Flourish - https://flourish.studio/

https://public.flourish.studio/visualisation/3166201/
Other Programs for Creating Data Visualizations

- Piktochart - https://piktochart.com/
- Canva - https://www.canva.com/
- Flourish - https://flourish.studio/
- Tableau - https://public.tableau.com/app/discover

https://public.tableau.com/views/CSLevelMarriageRates2000_1/MapBarChart?:language=en-US&:display_count=n&:origin=viz_share_link
Other Programs for Creating Data Visualizations

Piktochart - https://piktochart.com/
Canva - https://www.canva.com/
Flourish - https://flourish.studio/
Tableau - https://public.tableau.com/app/discover
ArcGIS
Stata
R
What are some – outlets for data visualization publication?
Outlets for “Publication”

- Public sites in which others can view your work
  - Flourish - https://flourish.studio/
  - Tableau - https://public.tableau.com/app/discover
Outlets for “Publication”

- Public sites in which others can view your work
  - Flourish - https://flourish.studio/
  - Tableau - https://public.tableau.com/app/discover
- Socius – Peer reviewed
  
https://journals.sagepub.com/doi/full/10.1177/23780231221090192

<table>
<thead>
<tr>
<th>State</th>
<th>Marriages, 2020</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observed</td>
</tr>
<tr>
<td>HI</td>
<td>10,444</td>
</tr>
<tr>
<td>NV</td>
<td>82,911</td>
</tr>
<tr>
<td>FL</td>
<td>92,979</td>
</tr>
<tr>
<td>NJ</td>
<td>88,342</td>
</tr>
<tr>
<td>PA</td>
<td>58,750</td>
</tr>
<tr>
<td>OR</td>
<td>23,556</td>
</tr>
<tr>
<td>GA</td>
<td>55,045</td>
</tr>
<tr>
<td>AK</td>
<td>8,172</td>
</tr>
<tr>
<td>SC</td>
<td>86,680</td>
</tr>
<tr>
<td>OH</td>
<td>56,000</td>
</tr>
<tr>
<td>CO</td>
<td>38,741</td>
</tr>
<tr>
<td>NH</td>
<td>8,358</td>
</tr>
<tr>
<td>ME</td>
<td>9,172</td>
</tr>
<tr>
<td>IA</td>
<td>13,305</td>
</tr>
<tr>
<td>MS</td>
<td>16,815</td>
</tr>
<tr>
<td>OK</td>
<td>23,307</td>
</tr>
<tr>
<td>AZ</td>
<td>56,870</td>
</tr>
<tr>
<td>WY</td>
<td>5,018</td>
</tr>
<tr>
<td>ND</td>
<td>98,373</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>State</th>
<th>Divorces, 2020</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observed</td>
</tr>
<tr>
<td>SC</td>
<td>20,924</td>
</tr>
<tr>
<td>GA</td>
<td>4,241</td>
</tr>
<tr>
<td>NH</td>
<td>43,986</td>
</tr>
<tr>
<td>*NY</td>
<td>9,763</td>
</tr>
<tr>
<td>OR</td>
<td>2,452</td>
</tr>
<tr>
<td>RI</td>
<td>2,604</td>
</tr>
<tr>
<td>*CT</td>
<td>9,164</td>
</tr>
<tr>
<td>*MI</td>
<td>30,794</td>
</tr>
<tr>
<td>OK</td>
<td>13,716</td>
</tr>
<tr>
<td>*TX</td>
<td>109,887</td>
</tr>
<tr>
<td>FL</td>
<td>62,755</td>
</tr>
<tr>
<td>IL</td>
<td>23,012</td>
</tr>
<tr>
<td>*TN</td>
<td>24,558</td>
</tr>
<tr>
<td>CO</td>
<td>16,877</td>
</tr>
<tr>
<td>*DE</td>
<td>2,842</td>
</tr>
<tr>
<td>*VT</td>
<td>2,094</td>
</tr>
<tr>
<td>WV</td>
<td>7,456</td>
</tr>
<tr>
<td>PA</td>
<td>28,684</td>
</tr>
<tr>
<td>*CA</td>
<td>103,457</td>
</tr>
<tr>
<td>*NV</td>
<td>10,016</td>
</tr>
<tr>
<td>AR</td>
<td>19,179</td>
</tr>
<tr>
<td>ME</td>
<td>4,961</td>
</tr>
<tr>
<td>*NM</td>
<td>13,334</td>
</tr>
<tr>
<td>OH</td>
<td>28,066</td>
</tr>
<tr>
<td>MO</td>
<td>15,091</td>
</tr>
<tr>
<td>IA</td>
<td>6,876</td>
</tr>
<tr>
<td>WI</td>
<td>15,932</td>
</tr>
<tr>
<td>*MN</td>
<td>13,821</td>
</tr>
<tr>
<td>ND</td>
<td>1,801</td>
</tr>
<tr>
<td>*HI</td>
<td>3,298</td>
</tr>
<tr>
<td>WY</td>
<td>2,218</td>
</tr>
<tr>
<td>AZ</td>
<td>21,980</td>
</tr>
<tr>
<td>WA</td>
<td>10,492</td>
</tr>
</tbody>
</table>
Outlets for “Publication”

- Public sites in which others can view your work
  - Flourish - https://flourish.studio/
  - Tableau - https://public.tableau.com/app/discover
- Socius – Peer reviewed
- NCFMR
Outlets for “Publication”

- Public sites in which others can view your work
  - Flourish - https://flourish.studio/
  - Tableau - https://public.tableau.com/app/discover
- Socius – Peer reviewed
- NCFMR
- Conference poster presentations
Do you —

always need a chart?
7. Do you need a chart?

$117 million

Increase in annual HMI spending in the U.S. from 2000 – 2010

Sources: U.S. Census Bureau, American Community Survey, 2008-2011; HMI spending data– Hawkins et al., 2013.
7. Do you need a chart?

\[
\begin{align*}
    2,298,977 & : 1,110,579 \\
    \text{NUMBER of MARRIAGES} & : \text{NUMBER of DIVORCES}
\end{align*}
\]

2 Marriages per every 1 Divorce

Source: U.S. Census Bureau, American Community Survey, 2015 1-yr est.
7. Do you need a chart?

Geographic Variation of Women’s Adjusted Marriage Rate Among States, 2020

Source: NCFMR analyses of U.S. Census Bureau, American Community Survey, 1-year Experimental PUMS, 2020
Other Helpful Sites

https://datavizproject.com/
http://circos.ca/
https://www.rapidtables.com/convert/color/hex-to-rgb.html
https://www.rawgraphs.io/
Stata
  • https://www.trentonmize.com/teaching/dmv
Colors
  • https://www.colorhexa.com/00338d-to-ffffff
  • https://mycolor.space/
  • https://designsystem.digital.gov/design-tokens/color/overview/
Accessibility
  • http://colorsafe.co/
  • https://contrastchecker.com/
  • https://webaim.org/resources/contrastchecker/