CDS DS 594

Chapter Two: Choosing an Effective Visual

1. Simple Text
    a. When there is just a number or two to share, simple text can be a great way to communicate
    b. Solely using the number and a few supporting words to clearly make the point is important
    c. Beyond potentially being misleading, putting one or only a couple of numbers in the table or graph simply causes the numbers to lose some of their oomph
    d. Example:

Children with a "Traditional" Stay-at-Home Mother

% of children with a married stay-at-home mother with a working husband

41 — 1970
20 — 2012
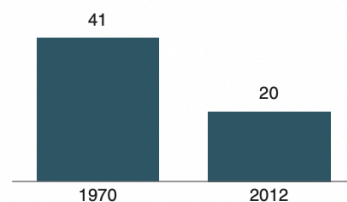
Note: Based on children younger than 18. Their mothers are categorized based on employment status in 1970 and 2012.

Source: Pew Research Center analysis of March Current Population Surveys Integrated Public Use Microdata Series (IPUMS-CPS), 1971 and 2013

Adapted from PEW RESEARCH CENTER

    e. The fact that there are some numbers does not mean that we need a graph
    f. The figure above doesn't do much to aid in the interpretation of the numbers
    g. Another Example:

20%
of children had a
traditional stay-at-home mom
in 2012, compared to 41% in 1970

    h. One consideration in the figure above is whether we want to show an entirely different metric
    i. In summary, when there is just a number or two that want we want to use for communication, use the numbers directly

2. Tables
    a. Tables interact with the verbal system, meaning that people read them
    b. Tables are great for communicating to a mixed audience whose members will each look for their particular row of interest
    c. When need to communicate multiple different units of measure, it is typically easier with a table than a graph
    d. Want the design to fade into the background, letting data take center stage and avoid heavy borders or shading compete for attention

Heavy borders

Group	Metric A	Metric B	Metric C
Group 1	$X.X	Y%	Z,ZZZ
Group 2	$X.X	Y%	Z,ZZZ
Group 3	$X.X	Y%	Z,ZZZ
Group 4	$X.X	Y%	Z,ZZZ
Group 5	$X.X	Y%	Z,ZZZ

Light borders

Group	Metric A	Metric B	Metric C
Group 1	$X.X	Y%	Z,ZZZ
Group 2	$X.X	Y%	Z,ZZZ
Group 3	$X.X	Y%	Z,ZZZ
Group 4	$X.X	Y%	Z,ZZZ
Group 5	$X.X	Y%	Z,ZZZ

Minimal borders

Group	Metric A	Metric B	Metric C
Group 1	$X.X	Y%	Z,ZZZ
Group 2	$X.X	Y%	Z,ZZZ
Group 3	$X.X	Y%	Z,ZZZ
Group 4	$X.X	Y%	Z,ZZZ
Group 5	$X.X	Y%	Z,ZZZ

3. Heatmap
    a. Mixing the detail we can include in a table while also making use of visual cues is via a heatmap
    b. Heatmap is way to visualize data in tabular format, where in place of the numbers, people leverage colored cells that convey the relative magnitude of the numbers
    c. Example:

Table

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

Heatmap

LOW-**HIGH**

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

    d. To reduce the mental processing, use color saturation to provide visual cues, helping the eyes and brains more quickly target the potential points of interest
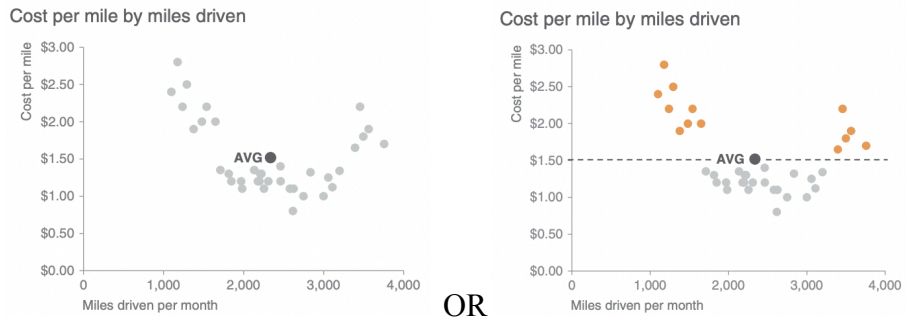4. Graphs
    a. Graphs interact with the visual systems, which is faster at processing information
    b. Types of graphs frequently used: points, lines, bars, and area
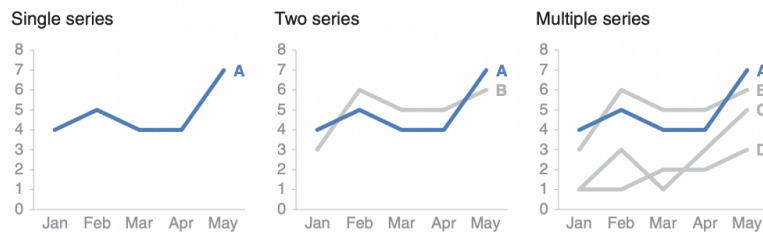5. Points
    a. Scatterplot
        i. Useful for showing the relationship between two things
        ii. More frequently used in scientific fields

iii.   Example:

OR

6. Lines
   a. Most commonly used to plot continuous data
   b. Since points are physically connected via the line, it implies a connection between the points that may not make sense for categorical data
   c. Continuous data, often, is in some unit of time: days, months, quarters, or years
   d. Example:

7. Slopegraph
   a. Useful when there are two time periods or points of comparison and want to quickly show relative increases and decreases or differences across various categories between two data points
   b. Best way to explain the value of and use case for slophegraphs is through a specific example
   c. They pack in a lot of information
   d. In addition to absolute values (points), the lines that connect them give visual increase or decrease in rate of change without ever having to explain that's what they are doing, or what exactly a "rate of change" is
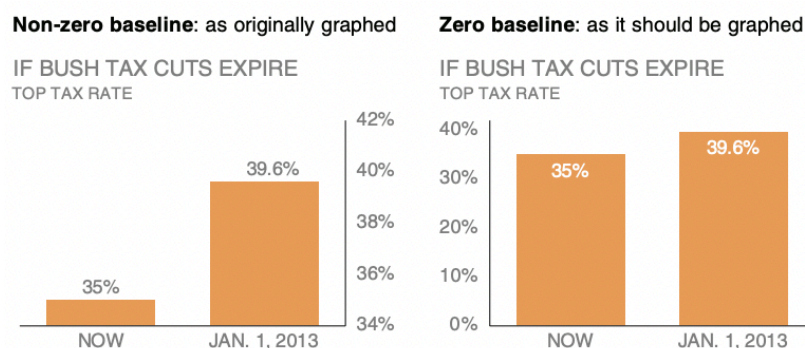   e. Example:

f. Whether a slopegraph will work in the specific situation depends on data itself

g. If many of the lines are overlapping, slopegraph may not work unless there is an emphasis:

Employee feedback over time

Survey category	Percent favorable
Peers	85% → 96%
Culture	80% → 91%
Work environment	76% → 75%
Leadership	59% → 62%
Career development	49% → 33%
Rewards & recognition	41% → 45%
Perf management	33% → 42%

2014   2015

Survey year

8. Bars

   a. Bar charts should be leveraged because they are common, as this means less of a learning curve for the audience

   b. Instead of using brain power to try to understand how to read the graph, the audience spends it figuring out what information to take away from the visual

   c. Bar charts easy for eyes to read

   d. It is easy to see quickly which category is the biggest, smallest, and the incremental difference between categories

   e. Important that bar charts always have a zero baseline (where the x-axis crosses the y-axis at zero) → otherwise, can get false visual comparison

   f. Example:

Non-zero baseline: as originally graphed

IF BUSH TAX CUTS EXPIRE
TOP TAX RATE

NOW 35%   JAN. 1, 2013 39.6%

Zero baseline: as it should be graphed

IF BUSH TAX CUTS EXPIRE
TOP TAX RATE

NOW 35%   JAN. 1, 2013 39.6%

   i. The right design is a better bar graph since it starts from 0

   g. Length of bars do not have a rule, but in general, the bars should be wider than the white space between the bars without going so wide
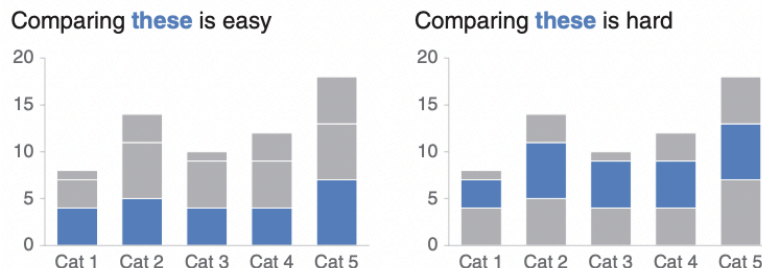
h. Example:
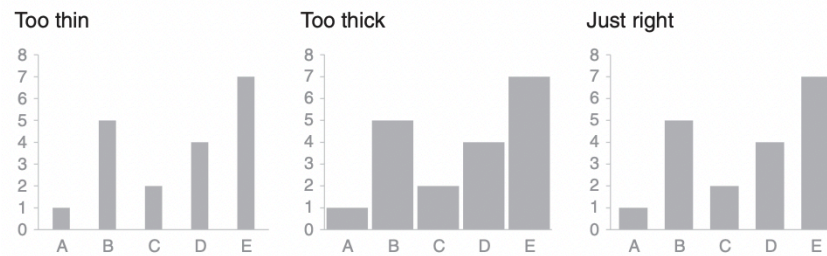
9. Vertical Bar Chart
    a. Plain vanilla bar chart is vertical bar chart or column chart
    b. Vertical bar charts can be single series, two series, or multiple series
    c. By adding more series of data, it becomes more difficult to focus on one at a time and pull out insight, so using multiple series bar charts requires caution
    d. Considering what the audience wants to compare is significant and structure the categorization hierarchy as easy as possible
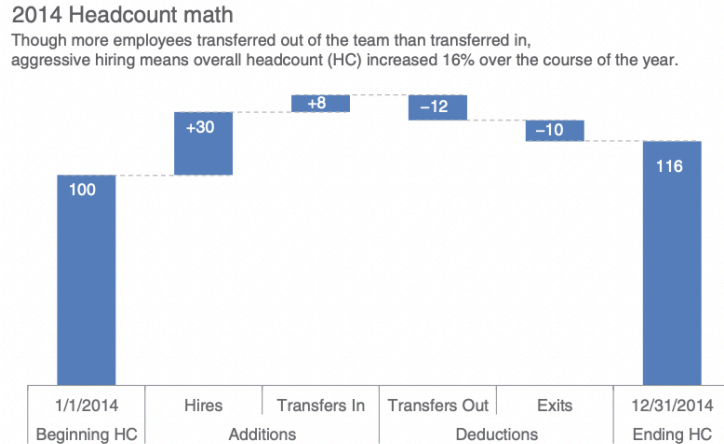   e. Example:
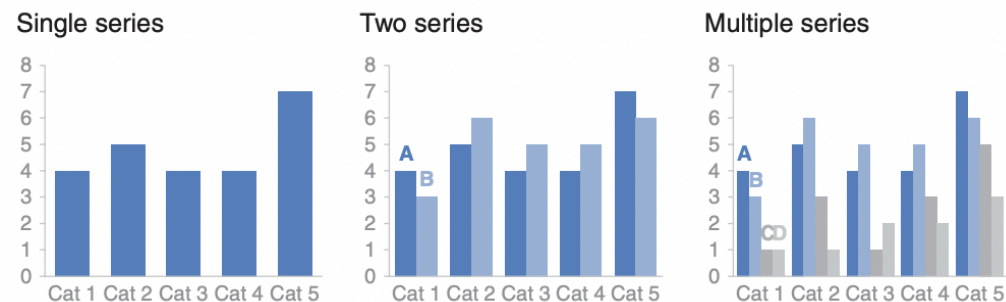
10. Stacked Vertical Bar Chart
    a. Use case for stacked vertical bar charts are more limited
    b. They are meant to allow to compare totals across categories and also see the subcomponent pieces within a given category
    c. It can quickly become visually overwhelming
    d. Hard to compare subcomponents across the various categories once people get beyond the bottom series because there is no longer a consistent baseline to use to compare
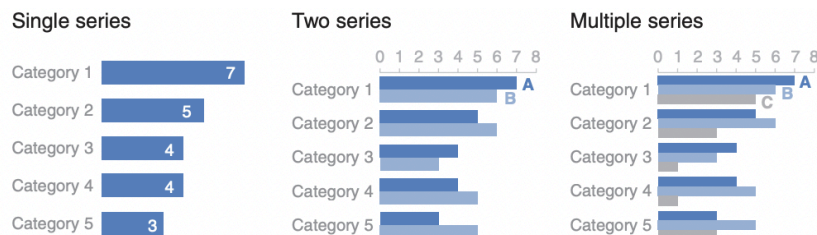   e. Example:

11. Waterfall Chart
   a. Can be used to pull apart the pieces of stacked bar chart to focus on one at a time, or to show a starting point, increases and decreases, and the resulting ending point
   b. Example:

2014 Headcount math
Though more employees transferred out of the team than transferred in, aggressive hiring means overall headcount (HC) increased 16% over the course of the year.

      i. On the left-hand side, we see what the employee headcount for the given team was at the beginning of the year
      ii. As we move to the right, we encounter the incremental additions: new hires and employees transferring into the team from other parts of the organization
      iii. This is followed by the deductions: transfers out of the tam to other parts of the organization and attrition
      iv. The final column represents employee headcount at the end of the year, after the additions and deductions have been applied to the beginning of year headcount
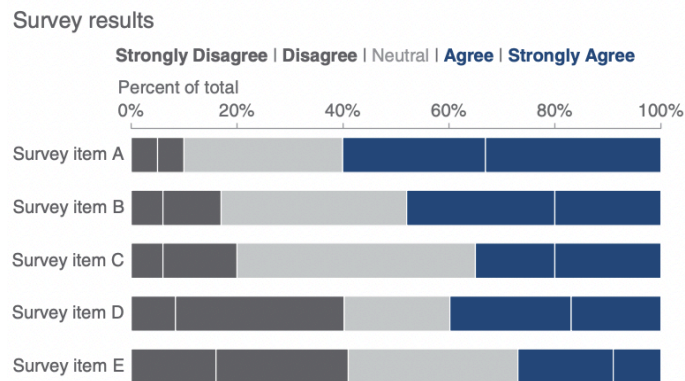12. Horizontal Bar Chart
   a. It is extremely easy to read
   b. Horizontal bar chart is especially useful if the category names are long, as the text is written from left to right, as most audiences read, making the graph legible for the audience
   c. The structure of the horizontal bar chart is such that the eyes hit the category names before the actual data, meaning that by the time people get to the data, they already know what it represents
   d. Example:

13. Stacked Horizontal Bar Chart
    a. Stacked horizontal bar charts can be used to show the totals across different categories but also give a sense of the subcomponent pieces
    b. Structured to show either absolute values or sum to 100%
    c. Example:

Survey results

Strongly Disagree | **Disagree** | Neutral | **Agree** | **Strongly Agree**

Percent of total

| | 0% | 20% | 40% | 60% | 80% | 100% |

Survey item A
Survey item B
Survey item C
Survey item D
Survey item E

14. Area
    a. Avoid most area graphs since humans' eyes do not do a great job of attributing quantitative value to two-dimensional space, which can render area graphs harder to read than some of the other types of visual displayed
    b. When need to visualize numbers of vastly different magnitudes, use area graphs
    c. Example:

Interview breakdown

Out of every **100** phone screens...

we bring **25** candidates onsite for interviews...
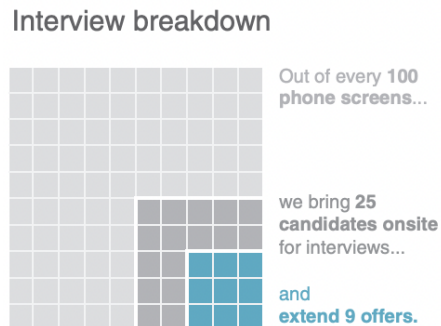
and **extend 9 offers.**

FIGURE 2.20   Square area graph

15. Other Types of Graphs
    a. Many other types of graphs out in the world
    b. When choosing a graph, selecting a graph type that will enable to clearly get the message across the audience is important
    c. With less familiar types of visuals, will likely need to take extra care in making them accessible and understandable

16. To be Avoided
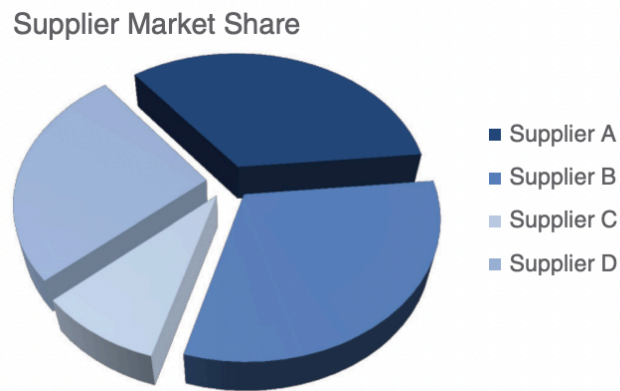   a. Pie Charts are Evil
        i. Example:

FIGURE 2.21   Pie chart

        ii. It is 3D and strange perspective that is been applied to the graph, tilting the pie and making the peices at the top appear farther away and thus smaller than they actually are
        iii. Human eye isn't good at ascribing quantitative value to two-dimensional space
        iv. Pie charts are hard for people to read
        v. When segments are close in size, it is difficult to tell which is bigger
        vi. When segments are not close in size, the best people can do is determine that one is bigger than the other, but cannot judge by how much
   vii. Even though people can add data labels, its visual is not worth the space it takes up
    b. Never Use 3D
        i. Only exception to use 3D is actually plotting a third dimension
        ii. Never use 3D to plot a single dimension
        iii. 3D skews the numbers, making them difficult to or impossible to interpret or compare
        iv. Adding 3D to graphs introduces unnecessary chart elements like side and floor panels
        v. Graphing applications do strange things when it comes to plotting values in 3D
        vi. Example:
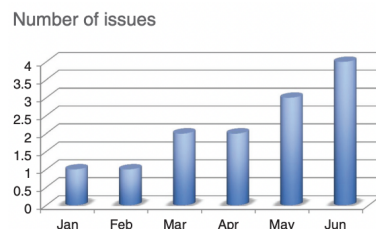
c. Secondary y-axis: Generally not a Good Idea
   i. Example:
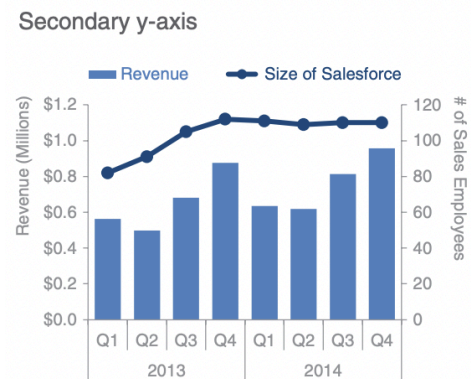
   ii. It takes some time and reading to understand which data should be read against which axis
   iii. Avoid using a secondary or right-hand y-axis
   iv. Instead, think about whether one of the following approaches will meet the needs
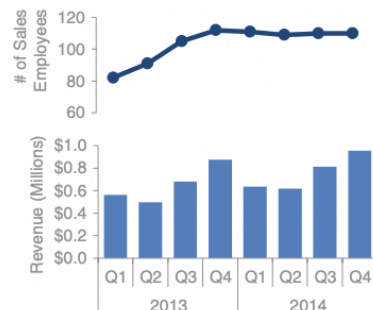   1. Don't show the second y-axis. Instead, label the data points that belong on the axis directly
      2. Pull the graphs apart vertically and have a separate y-axis for each but leverage the same x-axis across both
   v. Alternative Example:
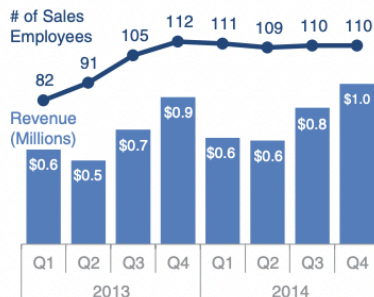
   vi. A third potential opinion is to link the axis to the data to be read against it through the use of color
   vii. When displaying two datasets against the same axis, it can imply a relationship that may or may not exist