

Visualizing Amounts

HUGEN 2073

Genomic Data Visualization and Integration

Slides borrowed/adapted from H.J. Park with permission

Today we'll discuss

- Tables
- Barplots
 - Grouped
 - Stacked
- Dotplots
- Line graphs
- Bubble charts

A table shows *detailed* information for a *limited* number of samples

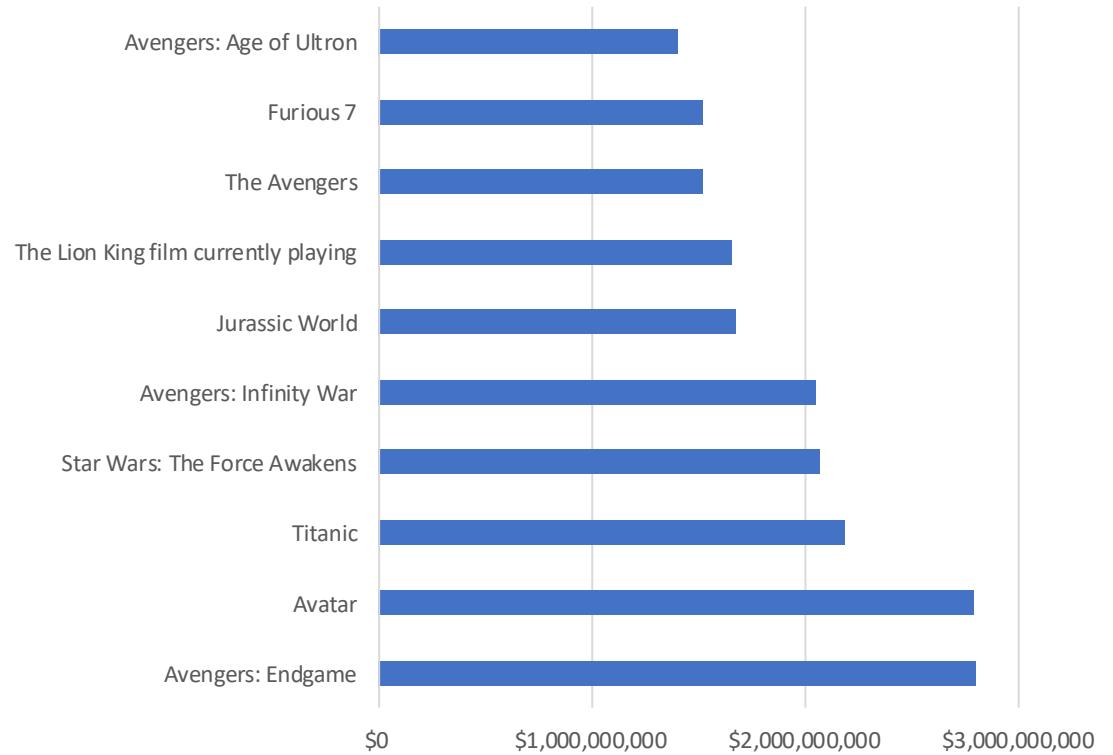
- Describe experiment or clinical data, thus important for data-oriented papers (clinical journals)
- Row and column are usually *not* interchangeable
- Row: an observation (e.g., a sample)
- Column: feature of the data

Highest-grossing films			
Rank	Title	Gross income	Year
1	Avengers: Endgame	\$2,797,800,564	2019
2	Avatar	\$2,789,679,794	2009
3	Titanic	\$2,187,463,944	1997
4	Star Wars: The Force Awakens	\$2,068,223,624	2015
5	Avengers: Infinity War	\$2,048,359,754	2018
6	Jurassic World	\$1,671,713,208	2015
7	The Lion King film currently playing	\$1,654,637,574	2019
8	The Avengers	\$1,518,812,988	2012
9	Furious 7	\$1,516,045,911	2015
10	Avengers: Age of Ultron	\$1,405,403,694	2015

https://en.wikipedia.org/wiki/List_of_highest-grossing_films

Tables can show exact values

Highest-grossing films			
Rank	Title	Gross income	Year
1	Avengers: Endgame	\$2,797,800,564	2019
2	Avatar	\$2,789,679,794	2009
3	Titanic	\$2,187,463,944	1997
4	Star Wars: The Force Awakens	\$2,068,223,624	2015
5	Avengers: Infinity War	\$2,048,359,754	2018
6	Jurassic World	\$1,671,713,208	2015
7	The Lion King film	\$1,654,637,574	2019
8	The Avengers	\$1,518,812,988	2012
9	Furious 7	\$1,516,045,911	2015
10	Avengers: Age of Ultron	\$1,405,403,694	2015

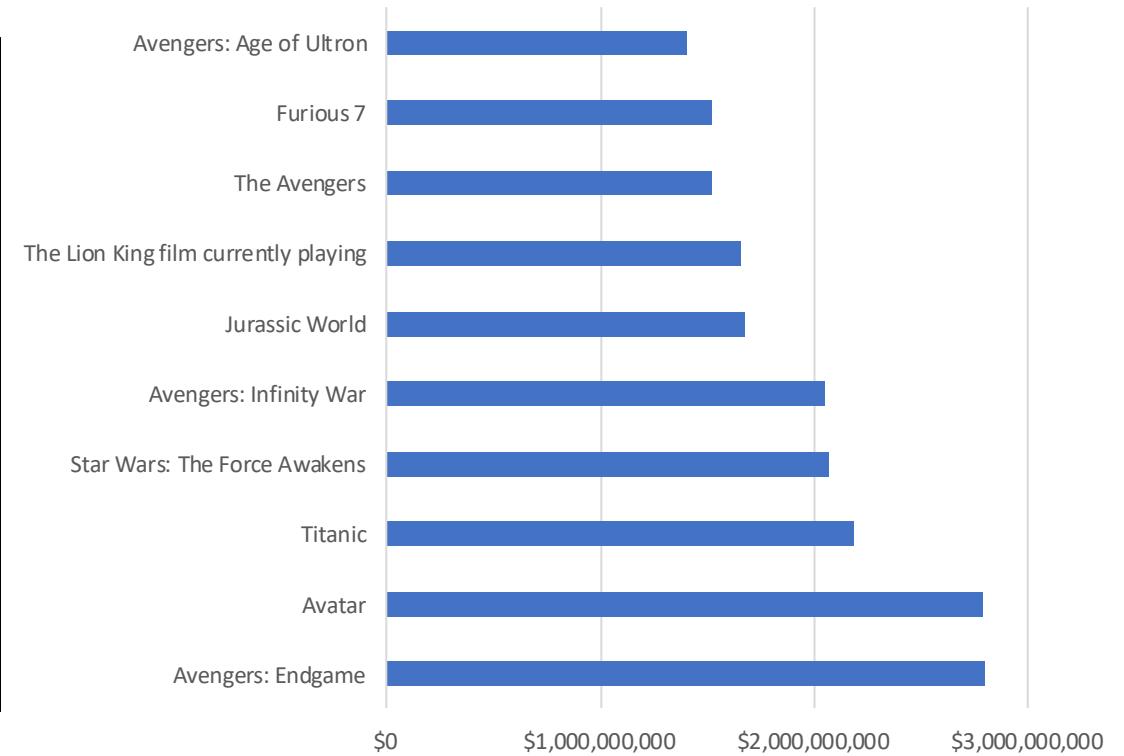


(There *is* a way a barchart could show exact values, but a table does so more naturally.)

Tables can show multiple variables

- E.g., if we're interested in ranking, gross income, *and* year
- An effective way to avoid 3D plotting

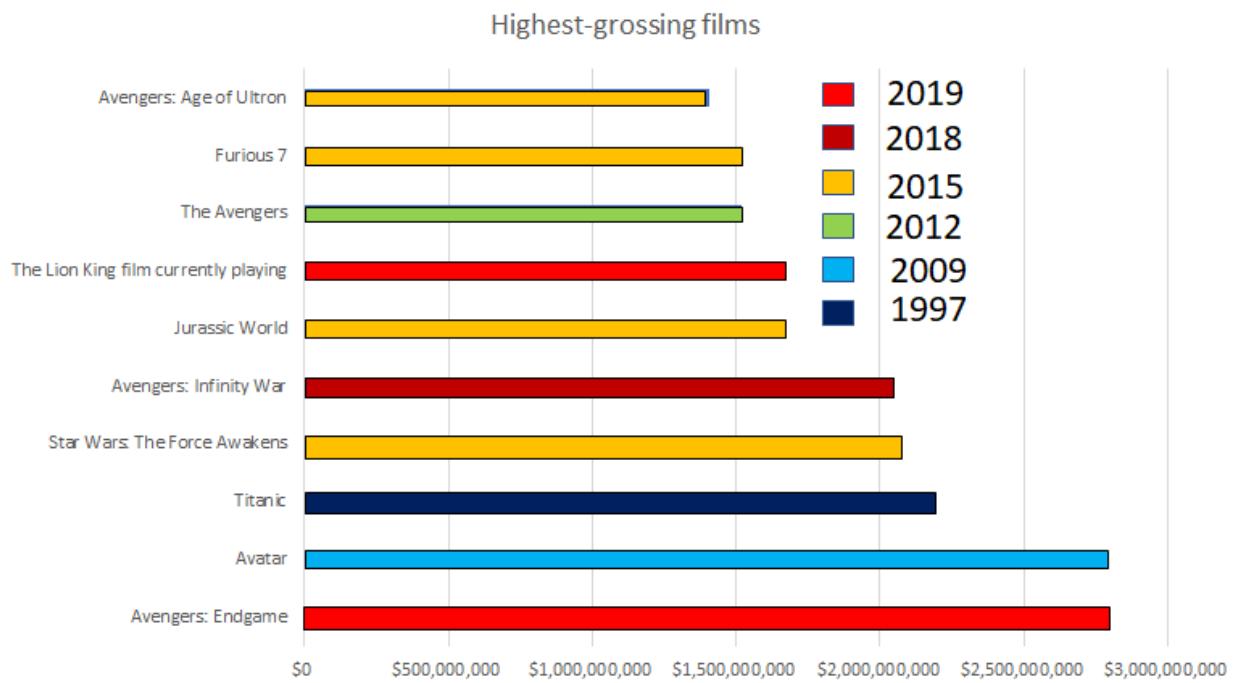
Highest-grossing films			
Rank	Title	Gross income	Year
1	Avengers: Endgame	\$2,797,800,564	2019
2	Avatar	\$2,789,679,794	2009
3	Titanic	\$2,187,463,944	1997
4	Star Wars: The Force Awakens	\$2,068,223,624	2015
5	Avengers: Infinity War	\$2,048,359,754	2018
6	Jurassic World	\$1,671,713,208	2015
7	The Lion King film currently playing	\$1,654,637,574	2019
8	The Avengers	\$1,518,812,988	2012
9	Furious 7	\$1,516,045,911	2015
10	Avengers: Age of Ultron	\$1,405,403,694	2015



Tables when you have many categories

- E.g., grouping by year
- The table can't do this quite as well as the barplot
- Too many labels/categories in the legend?

Highest-grossing films			
Rank	Title	Gross income	Year
1	Avengers: Endgame	\$2,797,800,564	2019
2	Avatar	\$2,789,679,794	2009
3	Titanic	\$2,187,463,944	1997
4	Star Wars: The Force Awakens	\$2,068,223,624	2015
5	Avengers: Infinity War	\$2,048,359,754	2018
6	Jurassic World	\$1,671,713,208	2015
7	The Lion King film currently playing	\$1,654,637,574	2019
8	The Avengers	\$1,518,812,988	2012
9	Furious 7	\$1,516,045,911	2015
10	Avengers: Age of Ultron	\$1,405,403,694	2015



Good for exploratory work on a small amount of data?

What is a good table for presentation?

In terms of maximizing data-ink ratio...

Rank	Title	Gross income(\$)	Year
1	Avengers: Endgame	2,797,800,564	2019
2	Avatar	2,789,679,794	2009
3	Titanic	2,187,463,944	1997
4	Star Wars: The Force Awakens	2,068,223,624	2015
5	Avengers: Infinity War	2,048,359,754	2018



Rank	Title	Gross income(\$)	Year
1	Avengers: Endgame	2,797,800,564	2019
2	Avatar	2,789,679,794	2009
3	Titanic	2,187,463,944	1997
4	Star Wars: The Force Awakens	2,068,223,624	2015
5	Avengers: Infinity War	2,048,359,754	2018

Rank	Title	Gross income(\$)	Year
1	Avengers: Endgame	2,797,800,564	2019
2	Avatar	2,789,679,794	2009
3	Titanic	2,187,463,944	1997
4	Star Wars: The Force Awakens	2,068,223,624	2015
5	Avengers: Infinity War	2,048,359,754	2018



Rank	Title	Gross income(\$)	Year
1	Avengers: Endgame	2,797,800,564	2019
2	Avatar	2,789,679,794	2009
3	Titanic	2,187,463,944	1997
4	Star Wars: The Force Awakens	2,068,223,624	2015
5	Avengers: Infinity War	2,048,359,754	2018

What is a good table for presentation?

In terms of maximizing data-ink ratio...

Rank	Title	Gross income(\$)	Year
1	Avengers: Endgame	2,797,800,564	2019
2	Avatar	2,789,679,794	2009
3	Titanic	2,187,463,944	1997
4	Star Wars: The Force Awakens	2,068,223,624	2015
5	Avengers: Infinity War	2,048,359,754	2018



Rank	Title	Gross income(\$)	Year
1	Avengers: Endgame	2,797,800,564	2019
2	Avatar	2,789,679,794	2009
3	Titanic	2,187,463,944	1997
4	Star Wars: The Force Awakens	2,068,223,624	2015
5	Avengers: Infinity War	2,048,359,754	2018

Rank	Title	Gross income (\$)	Year
1	Avengers: Endgame	2,797,800,564	2019
2	Avatar	2,789,679,794	2009
3	Titanic	2,187,463,944	1997
4	Star Wars: The Force Awakens	2,068,223,624	2015
5	Avengers: Infinity War	2,048,359,754	2018



Rank	Title	Gross income (\$)	Year
1	Avengers: Endgame	2,797,800,564	2019
2	Avatar	2,789,679,794	2009
3	Titanic	2,187,463,944	1997
4	Star Wars: The Force Awakens	2,068,223,624	2015
5	Avengers: Infinity War	2,048,359,754	2018

What is a good table for presentation?

In terms of alignment of text/number/single character columns...

Rank	Title	Gross income	Year
1	Avengers: Endgame	2,797,800,564	2019
2	Avatar	2,789,679,794	2009
3	Titanic	2,187,463,944	1997
4	Star Wars: The Force Awakens	2,068,223,624	2015
5	Avengers: Infinity War	2,048,359,754	2018

Rank	Title	Gross income	Year
1	Avengers: Endgame	2,797,800,564	2019
2	Avatar	2,789,679,794	2009
3	Titanic	2,187,463,944	1997
4	Star Wars: The Force Awakens	2,068,223,624	2015
5	Avengers: Infinity War	2,048,359,754	2018

Rank	Title	Gross income (\$)	Year
1	Avengers: Endgame	2,797,800,564	2019
2	Avatar	2,789,679,794	2009
3	Titanic	2,187,463,944	1997
4	Star Wars: The Force Awakens	2,068,223,624	2015
5	Avengers: Infinity War	2,048,359,754	2018

Rank	Title	Gross income (\$)	Year
1	Avengers: Endgame	2,797,800,564	2019
2	Avatar	2,789,679,794	2009
3	Titanic	2,187,463,944	1997
4	Star Wars: The Force Awakens	2,068,223,624	2015
5	Avengers: Infinity War	2,048,359,754	2018

What is a good table for presentation?

In terms of decimal places... generally use only 1 or 2

	2010	2011	2012	2013	2014	2015	2016	2017
Australia	101.7176	105.8239	106.6434	107.7288	106.7534	108.2793	110.0521	112.6886
Cambodia	56.9635	94.6286	129.2907	134.8927	133.9288	134.3668	126.3484	116.0421
China	63.1734	72.1219	80.8723	88.8862	92.5173	92.4809	97.2521	104.5817
Fiji	81.1582	83.8440	98.3073	105.7622	98.9129	108.2723	116.2363	114.1814
Hong Kong, China	196.3458	216.4354	230.6025	237.4339	235.7260	230.8188	240.7962	249.0243
Indonesia	87.1213	101.6679	113.2915	124.2805	127.6139	131.2928	147.6640	173.8402
Japan	95.9046	103.3117	109.8912	115.2554	123.1687	125.4618	130.6103	133.4504
Korea (Rep. of)	102.4507	105.5526	107.3514	108.9928	113.7039	116.4871	120.6797	124.8645
Malaysia	120.4409	128.0291	141.6663	144.7652	148.6323	143.5530	139.3678	133.8798
Myanmar	1.1843	2.4600	7.3149	13.2801	55.9072	78.2268	95.6532	89.8458
New Zealand	107.7788	109.0901	110.1675	105.5002	111.6780	121.3558	124.4413	136.0019
Philippines	88.7156	98.8579	105.2771	104.4095	111.2123	115.8497	116.2376	110.3956
Singapore	145.5308	150.5849	153.0576	157.4027	148.7388	148.7391	150.4805	148.2402
Thailand	106.7216	114.6881	125.3052	137.7235	141.9184	149.9353	173.7771	176.0347
Viet Nam	126.1072	142.3556	145.5732	135.2335	147.1157	128.5904	127.5261	125.6177

Rule of thumb for good table presentation

Rank	Title	Worldwide gross	Year
1	Avengers: Endgame	\$2,797,800,564	2019
2	Avatar	\$2,789,679,794	2009
3	Titanic	\$2,187,463,944	1997
4	Star Wars: The Force Awakens	\$2,068,223,624	2015
5	Avengers: Infinity War	\$2,048,359,754	2018

1. A horizontal line between the header and the first data row only
2. Text columns should be left aligned
3. Number columns should be right aligned and should use the same number of decimal digits throughout. (Columns containing single characters are centered.)

Rule of thumb for good table presentation

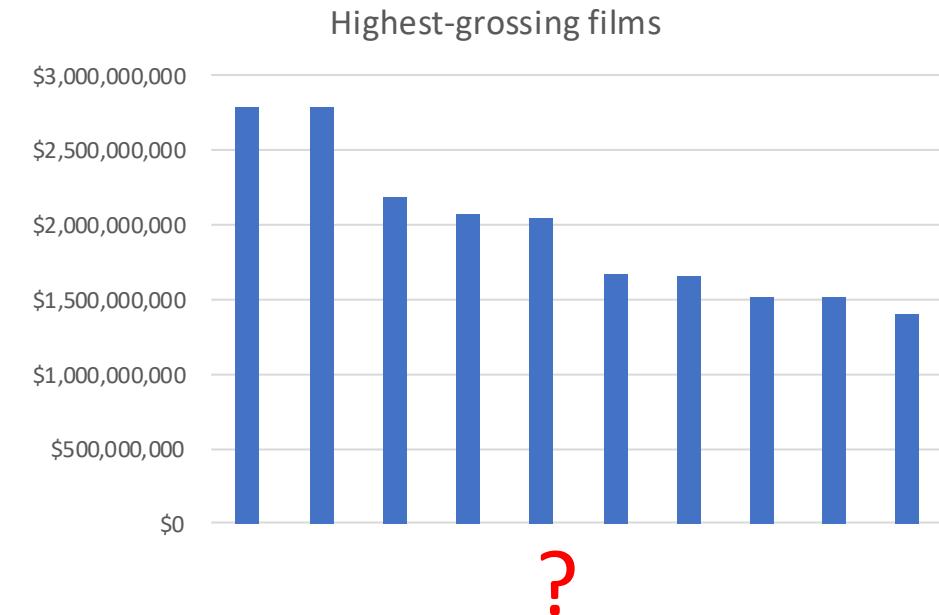
Rank	Title	Worldwide gross (\$)	Year
1	Avengers: Endgame	2,797,800,564	2019
2	Avatar	2,789,679,794	2009
3	Titanic	2,187,463,944	1997
4	Star Wars: The Force Awakens	2,068,223,624	2015
5	Avengers: Infinity War	2,048,359,754	2018

4. The header fields are aligned with their data
5. Unit should be notated in header
6. For big numbers, use commas (1,000 separator)

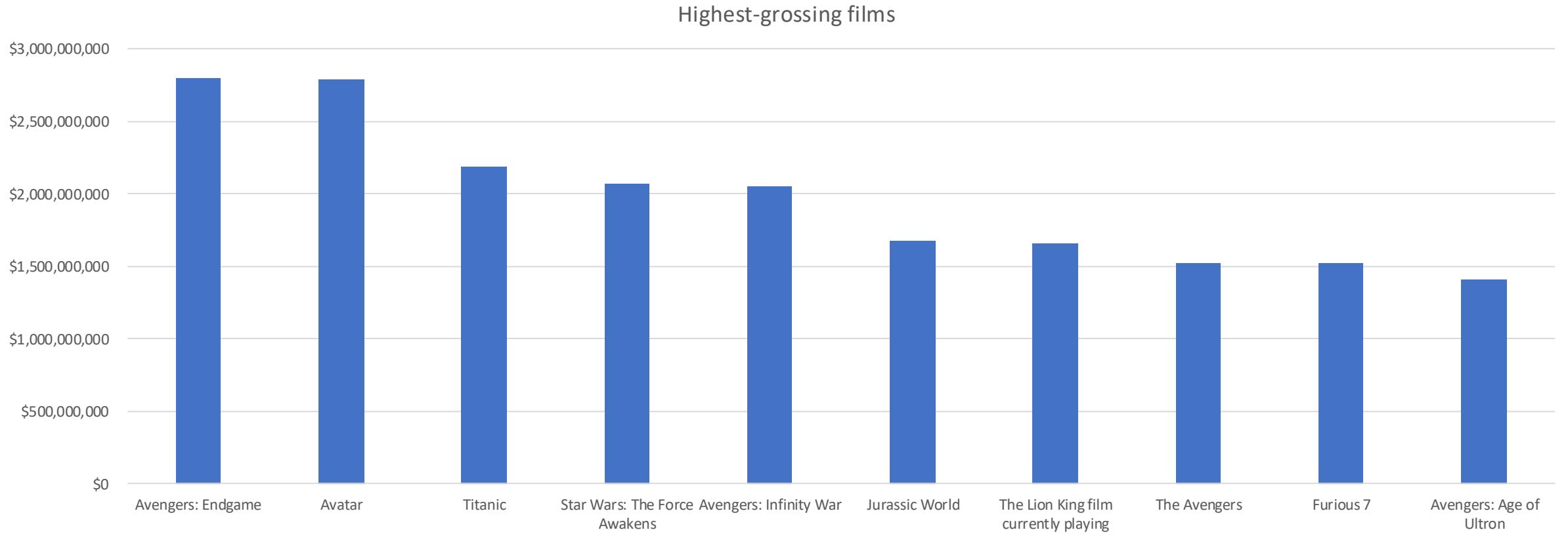
Barplot to replace table with a particular focus

Let's say we are interested in the gross income

Rank	Title	gross income (\$)	Year
1	Avengers: Endgame	2,797,800,564	2019
2	Avatar	2,789,679,794	2009
3	Titanic	2,187,463,944	1997
4	Star Wars: The Force Awakens	2,068,223,624	2015
5	Avengers: Infinity War	2,048,359,754	2018
6	Jurassic World	1,671,713,208	2015
7	The Lion King film currently playing	1,654,637,574	2019
8	The Avengers	1,518,812,988	2012
9	Furious 7	1,516,045,911	2015
10	Avengers: Age of Ultron	1,405,403,694	2015

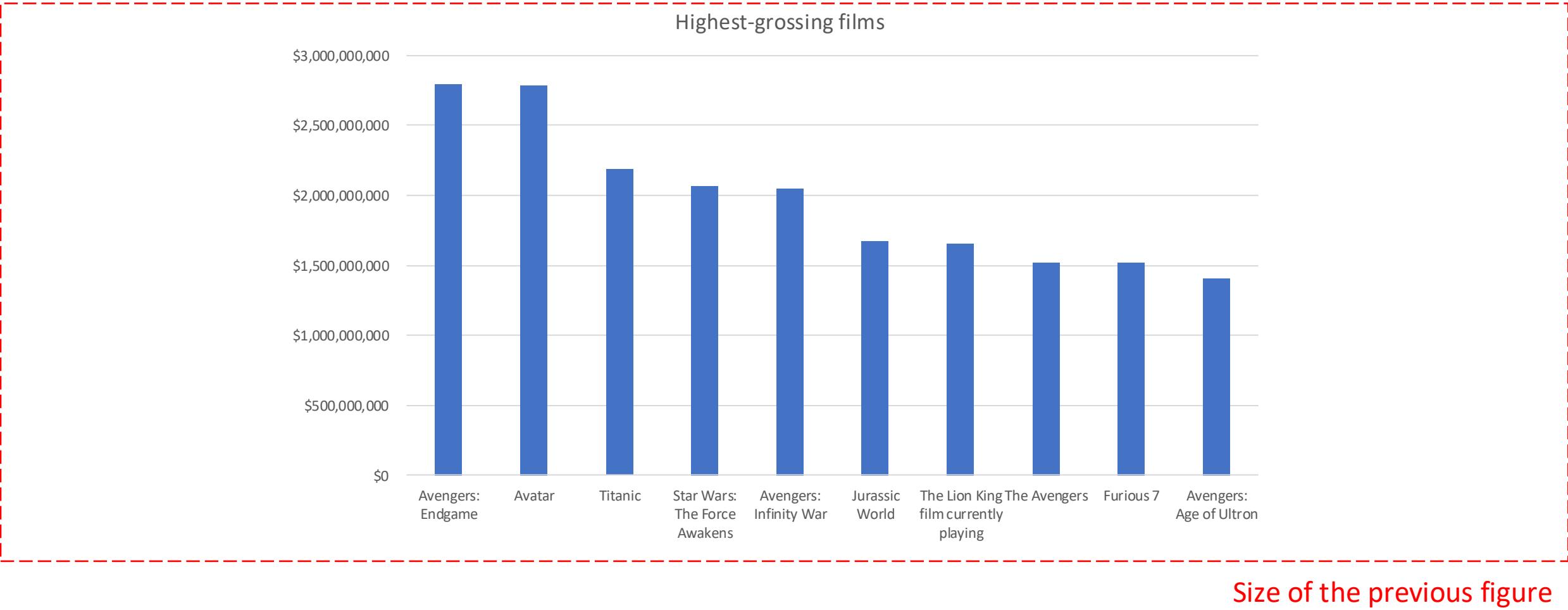


Long labels?

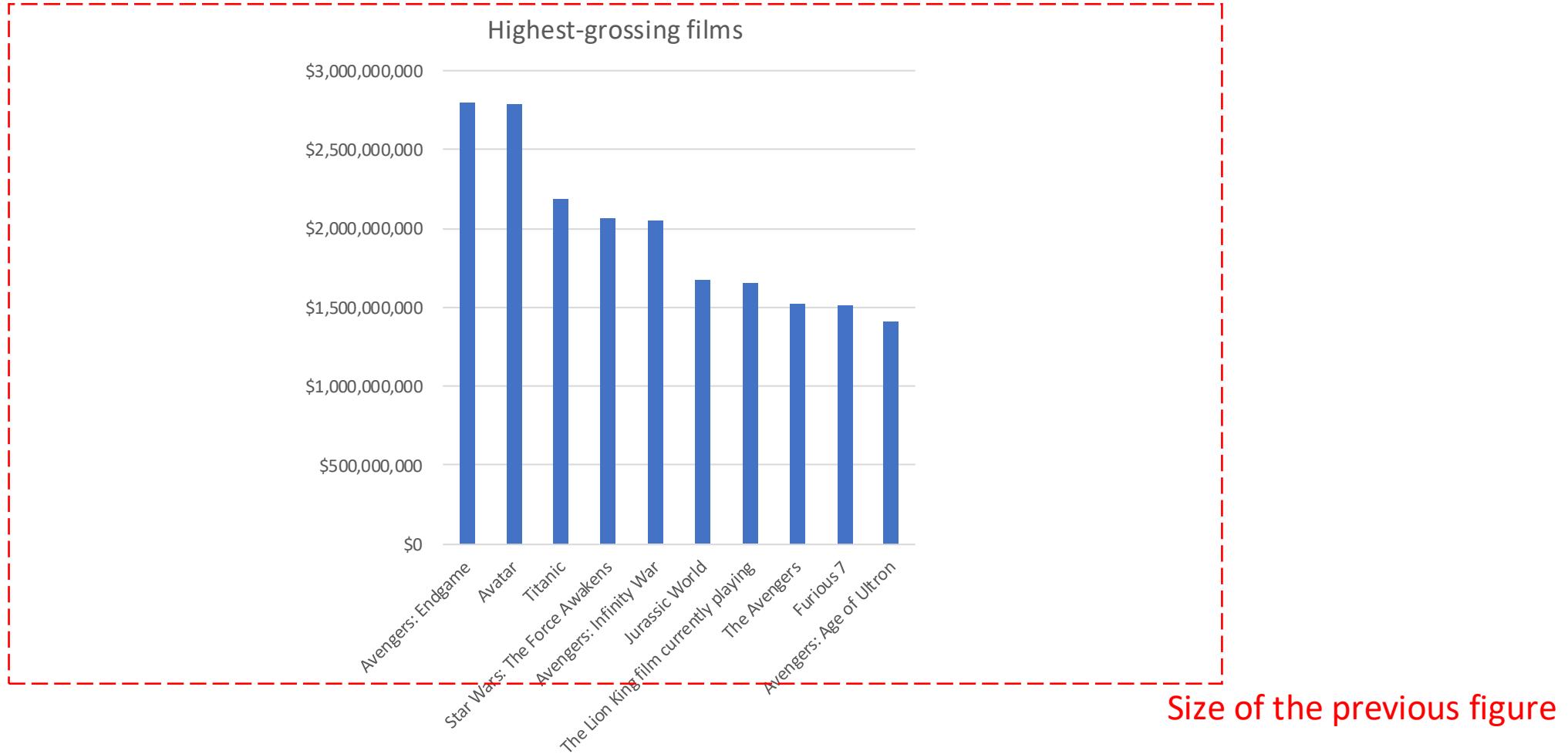


Maximizing not only data-ink ratio and readability, but also information amount per space

Multi-line labels?



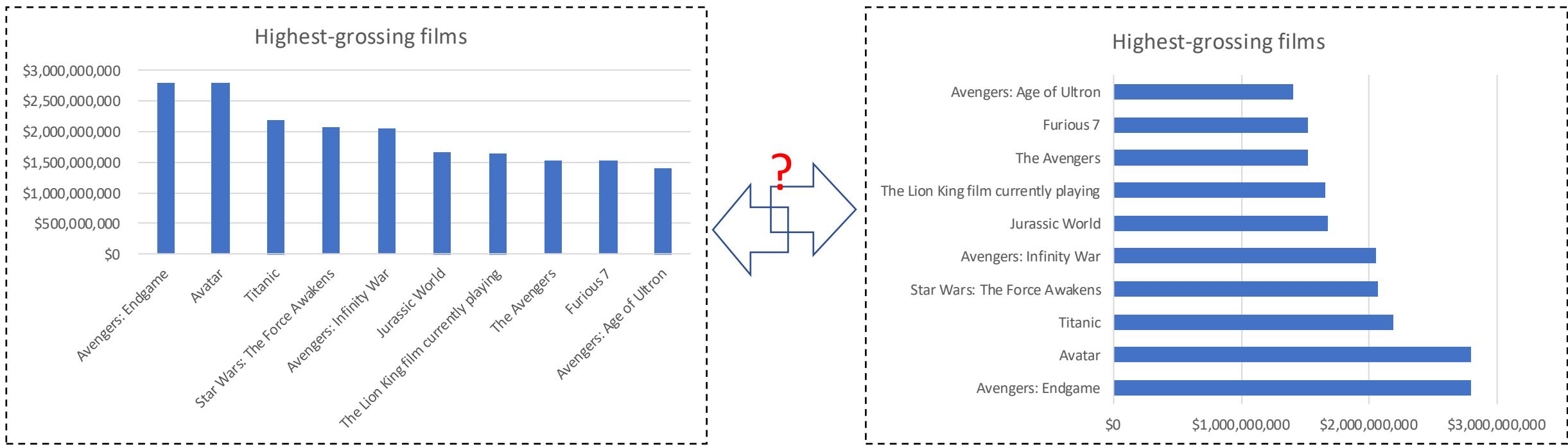
Tilted labels?



Efficient figures make information readily accessible

The same amount of information in

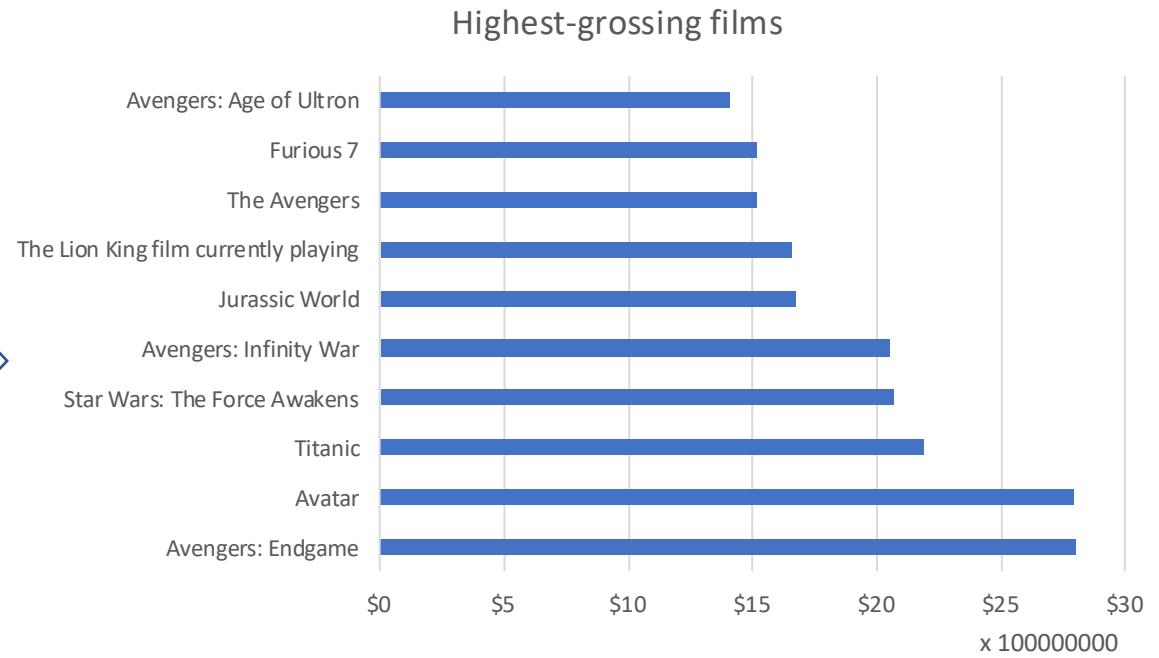
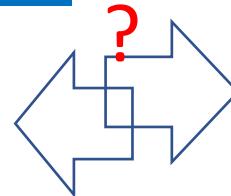
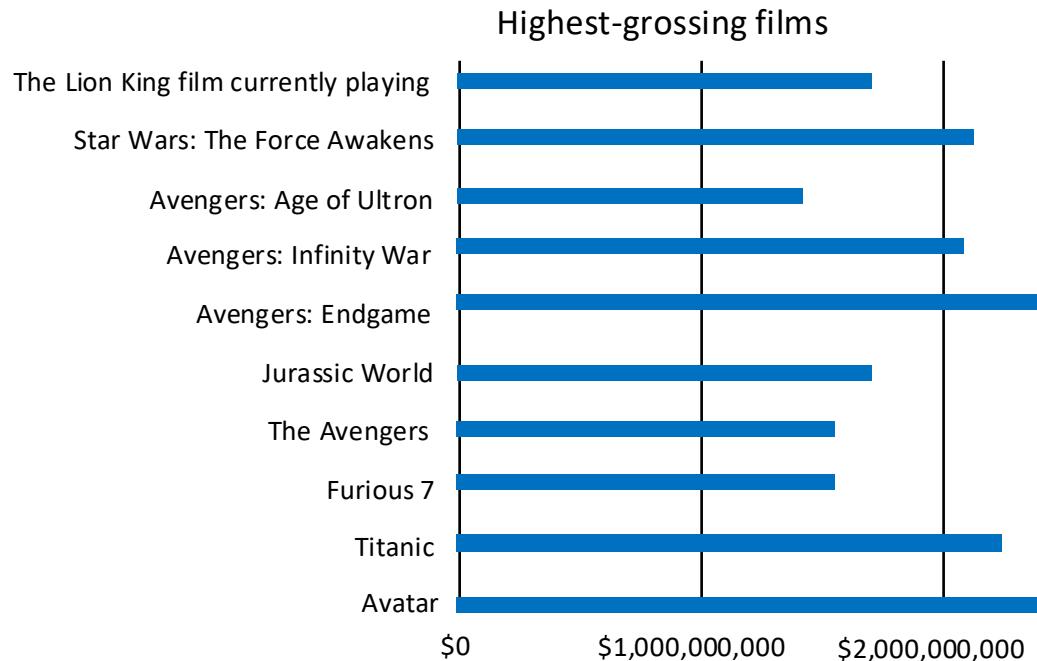
- Bar height and width
- Font size



Maximizing data-ink ratio, **readability**, and information amount

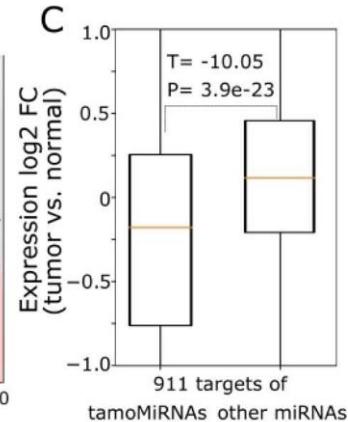
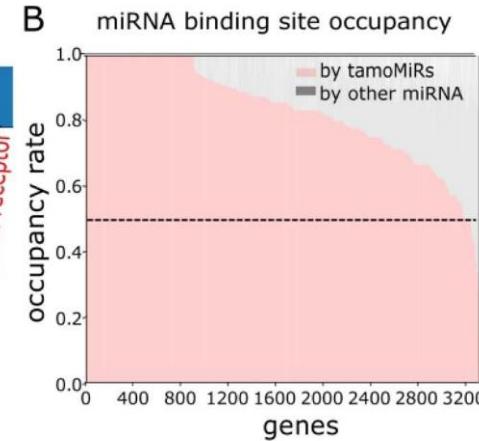
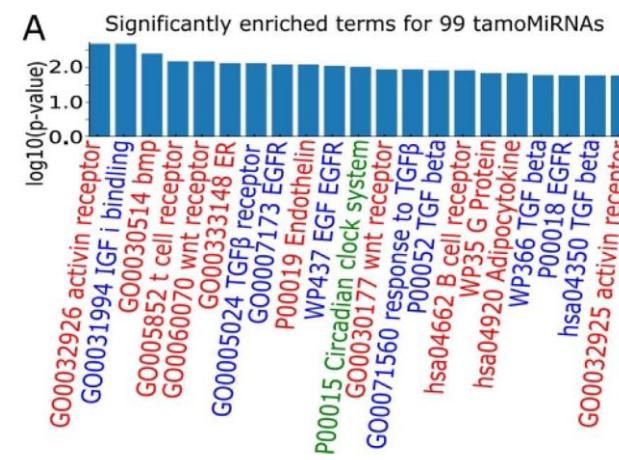
Sorting can improve efficiency

For readability, better to sort by bar height or by bar category text length? Why?

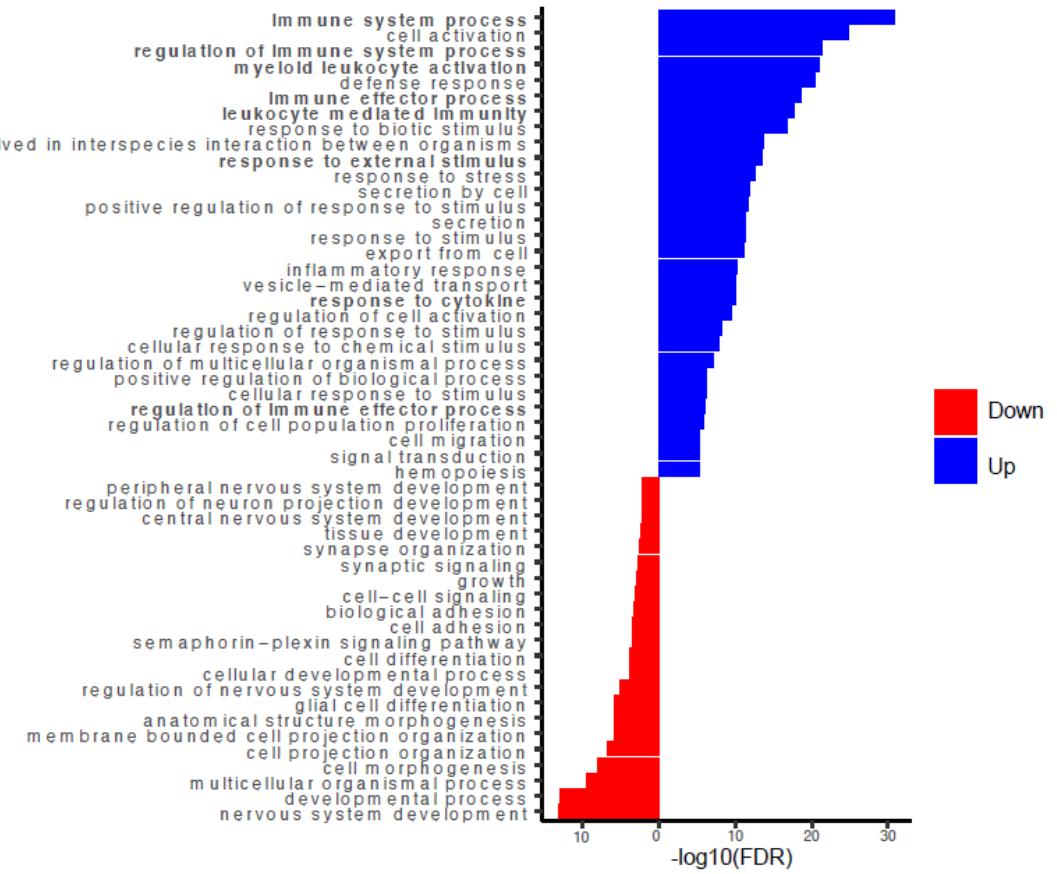


Other considerations in using barcharts

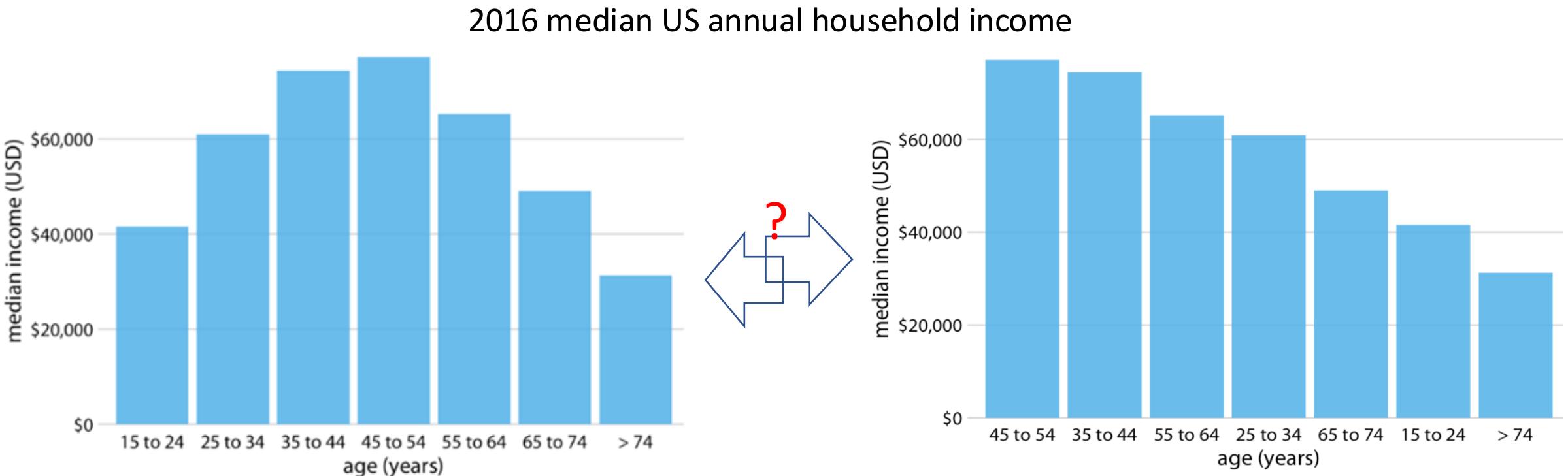
- Balance with other subfigures
- Bars of different directions (sort by sign, too)



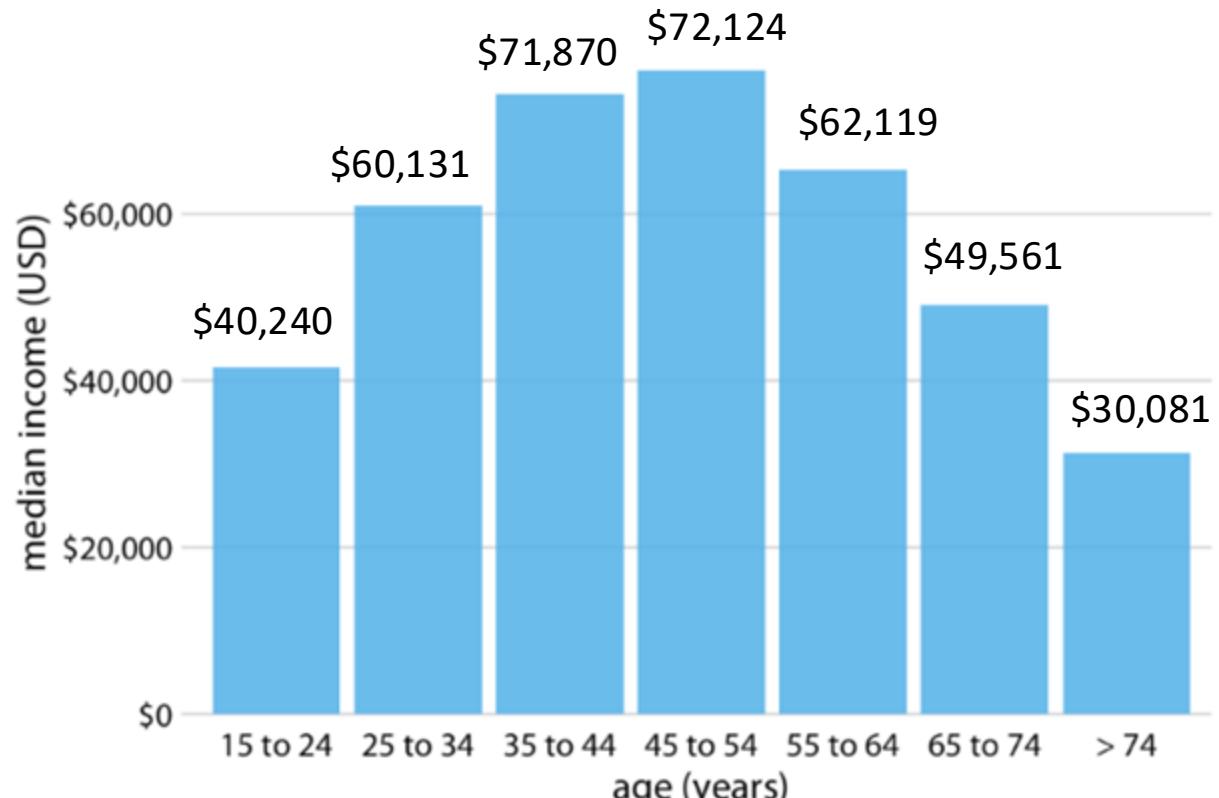
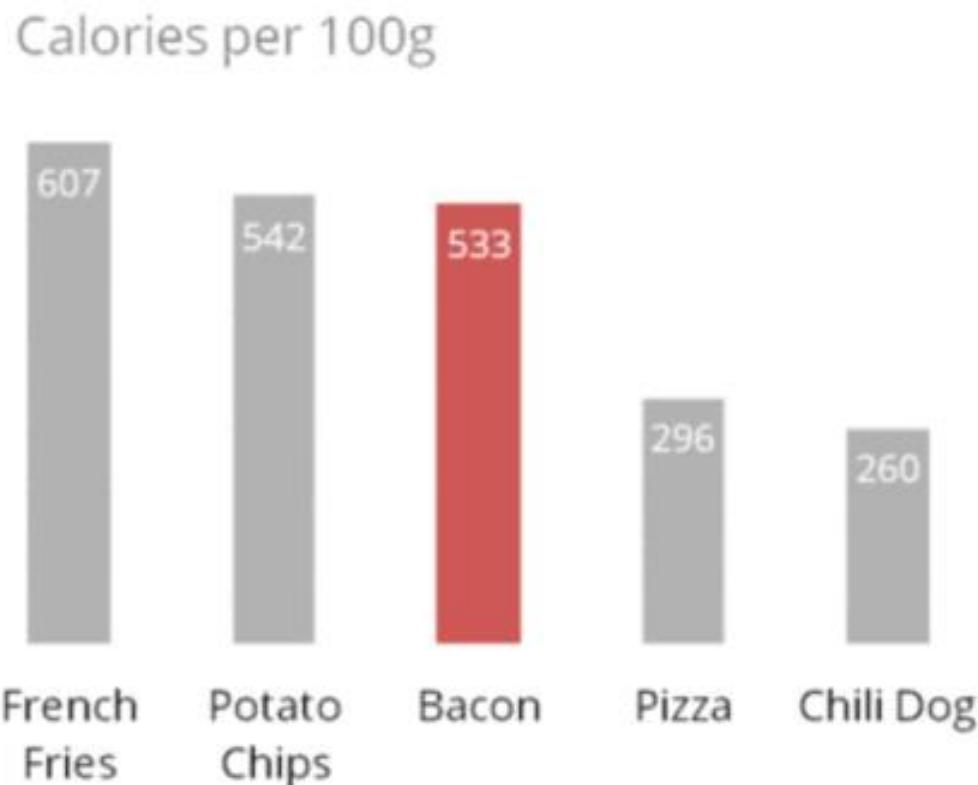
Kim et al. BIB 2020



Sorting can be more efficient (not always)



Barplot can visualize precise values

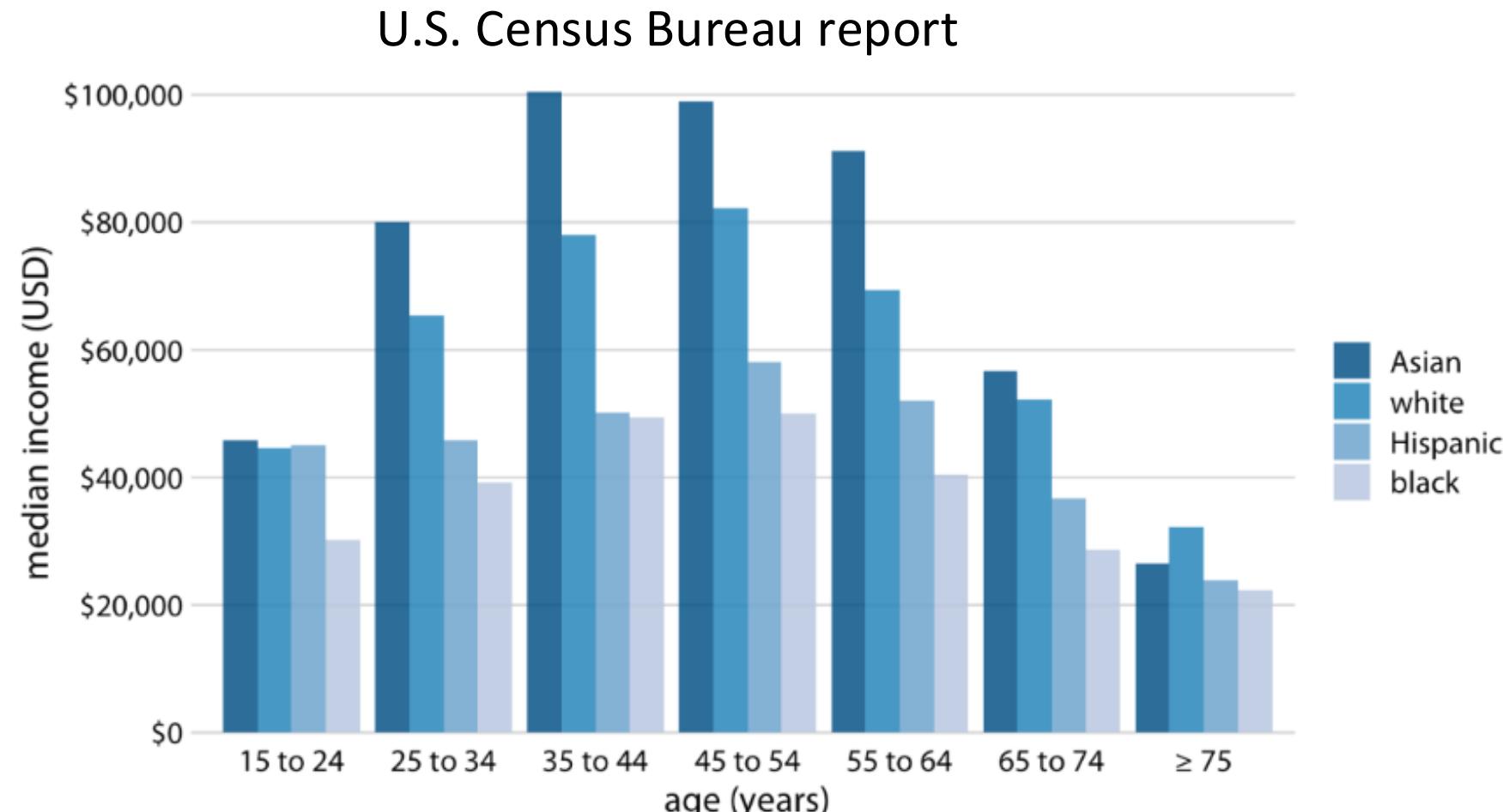


Adapted from Fundamentals of Data Visualization, Wilke, O'Reilly, 1st Ed.

- Barcharts can show exact values
- When to do this would be very context-specific

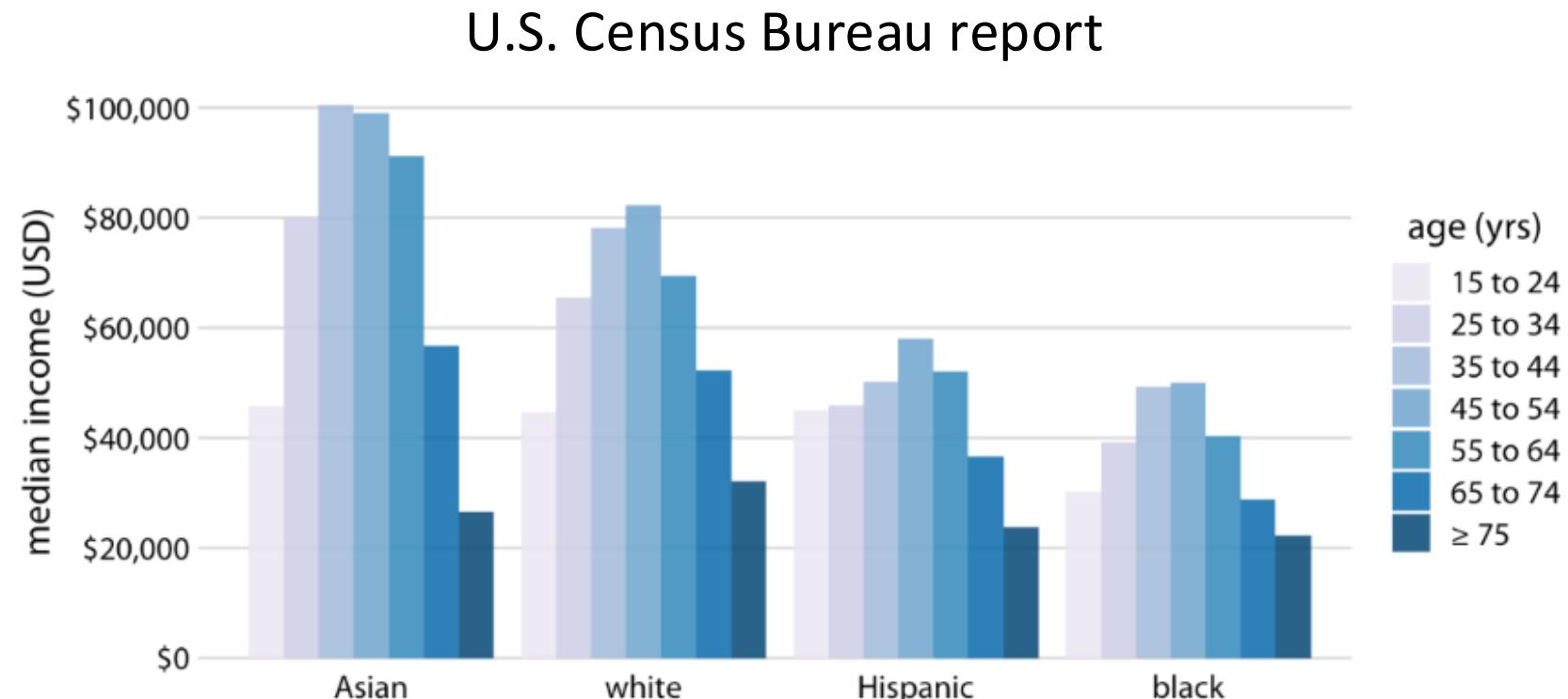
Grouped barplots show values in multiple categories

- A group of bars along the x-axis by one variable and then bars within each group by another variable
- Comparison among races is emphasized (not among age)
- Interested in certain age groups, too



Grouped barplot visualize amount in multiple variables

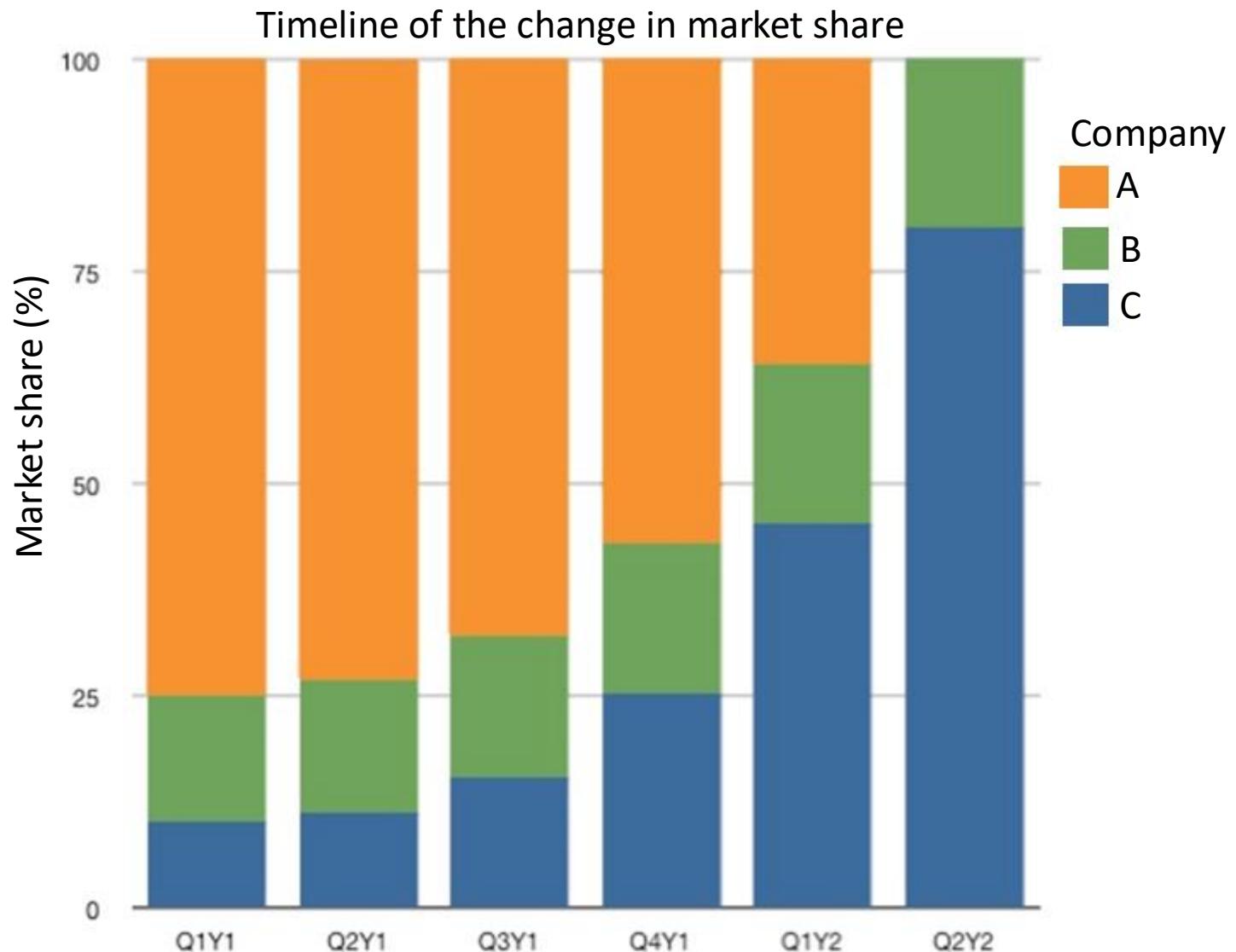
- Comparison among ages is emphasized
- If equally interesting, this may be better because
 - Sequential palette easier to read (for > 2 groups)
 - Comfortable seeing normal-like distribution (or other shapes that directly show statistical properties)



Stacked barplots visualize amount in multiple variables

Company	Q1Y1	Q2Y1	Q3Y1	Q4Y1	Q1Y2	Q2Y2
Blue	10	11	15	25	45	80
Green	15	16	17	18	19	20
Orange	75	73	68	57	36	0

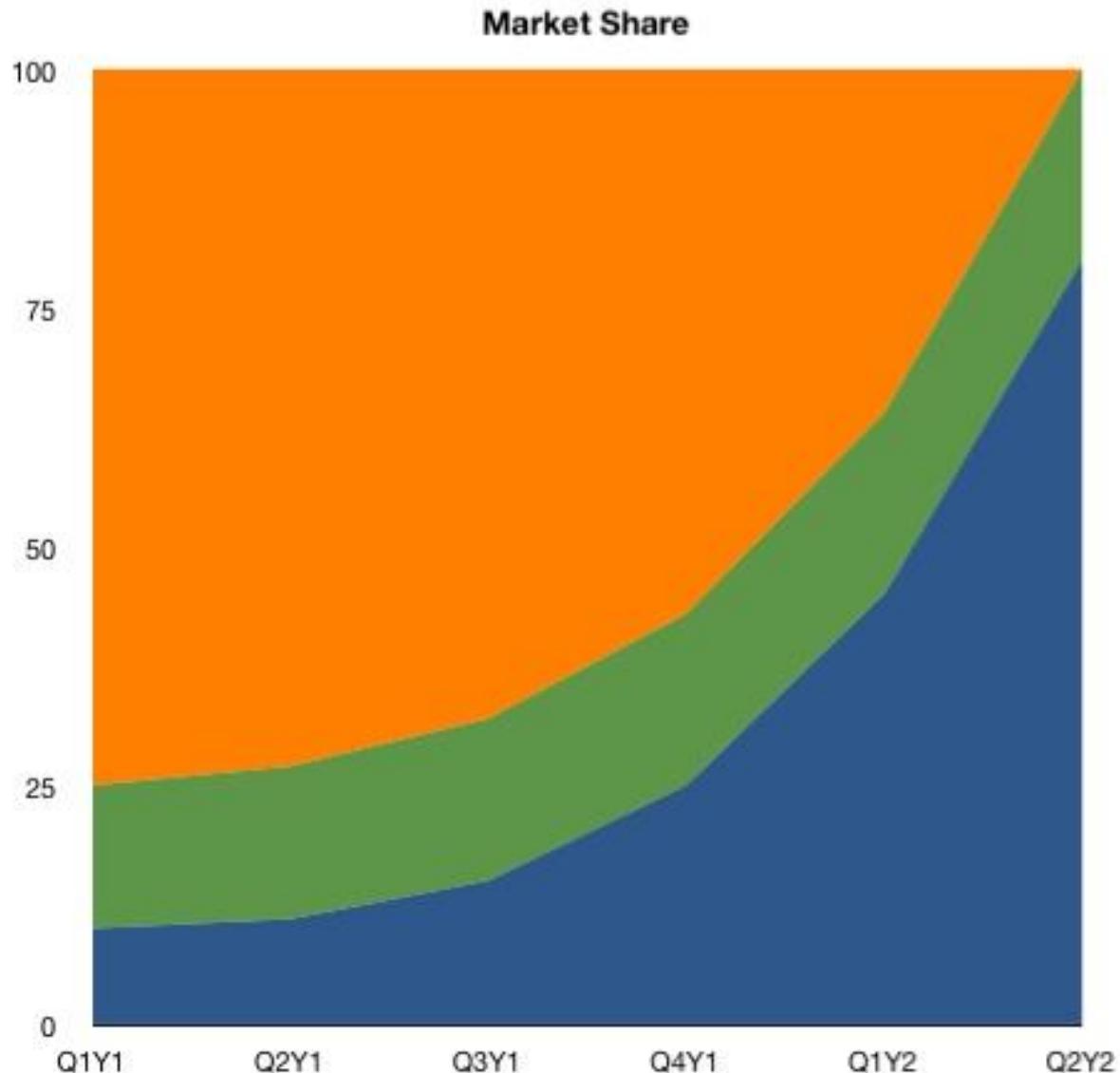
- Pros: preferable when
 - The total amount is fixed (as is always the case when % is presented)
 - The trend is somewhat clear
- Cons: not clear how each bar actually occupies!



Stacked area chart can distort the trend

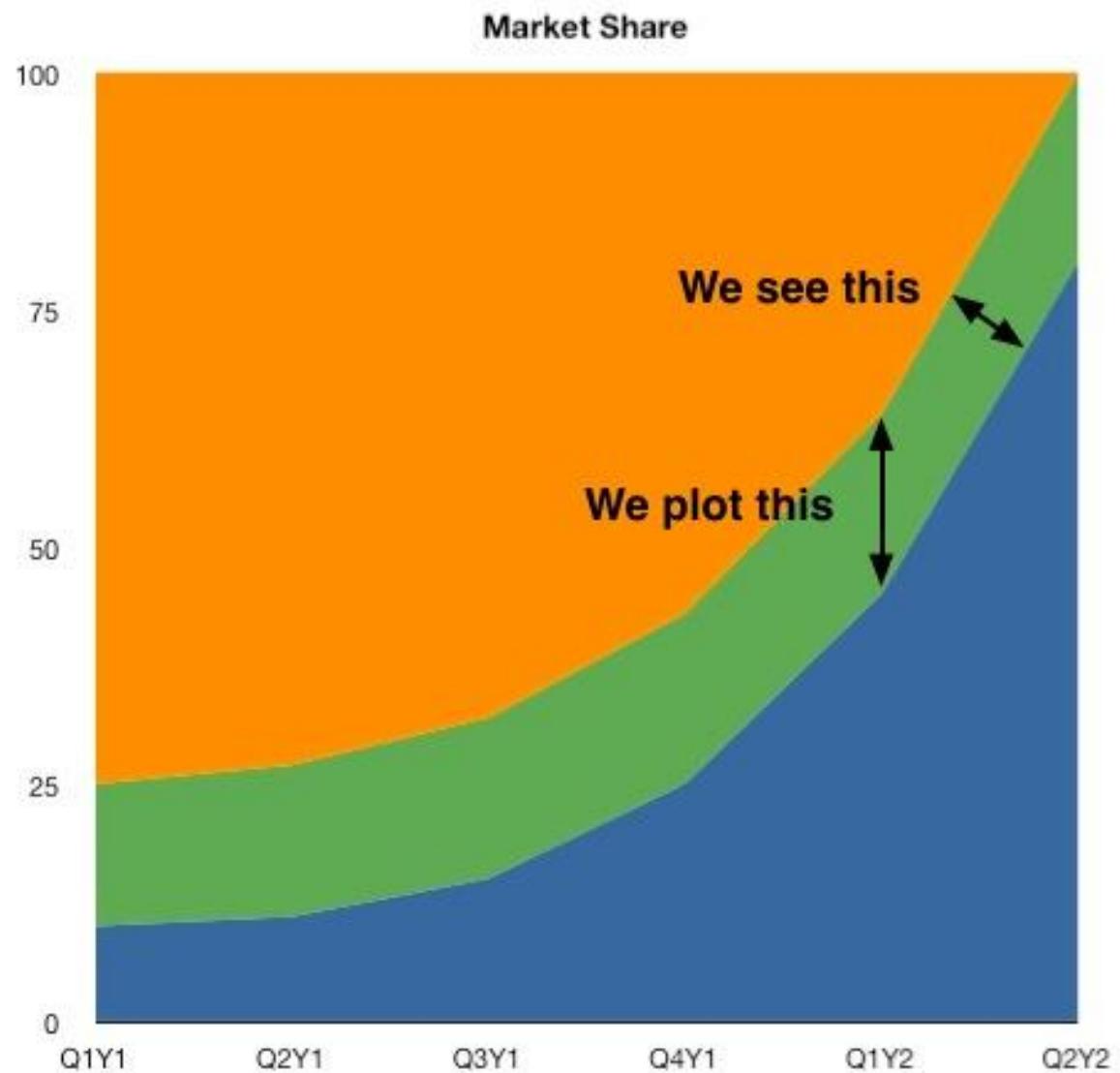
- General trend preserved
- Different from bars put together
- Question: How did Green do over this period?

Company	Q1Y1	Q2Y1	Q3Y1	Q4Y1	Q1Y2	Q2Y2
Blue	10	11	15	25	45	80
Green	15	16	17	18	19	20
Orange	75	73	68	57	36	0



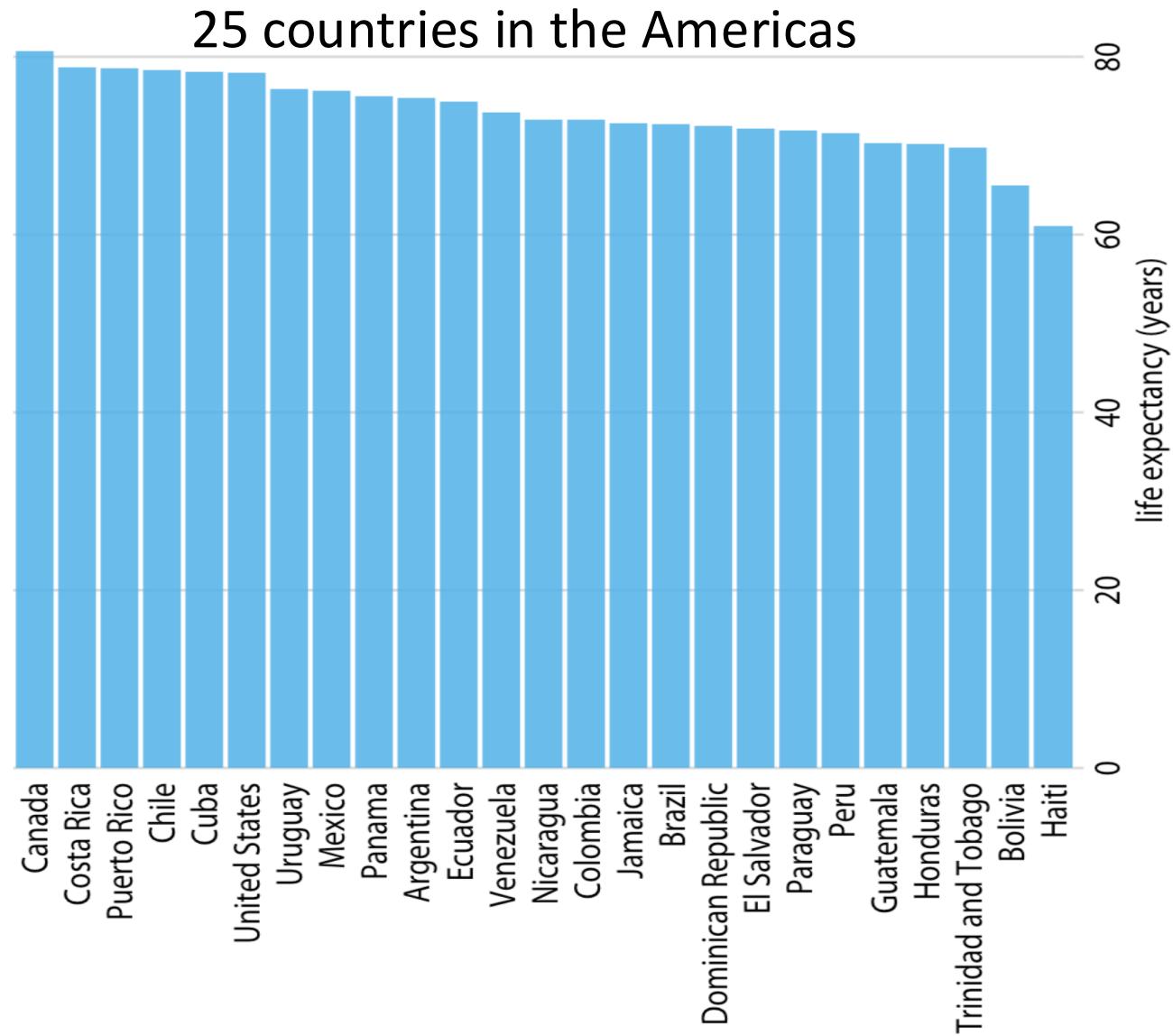
Stacked area chart can distort the trend

- Market share shown vertically
- But we may perceive thickness of a stream at right angles to that
- The overall trend we perceive is determined by “major players”



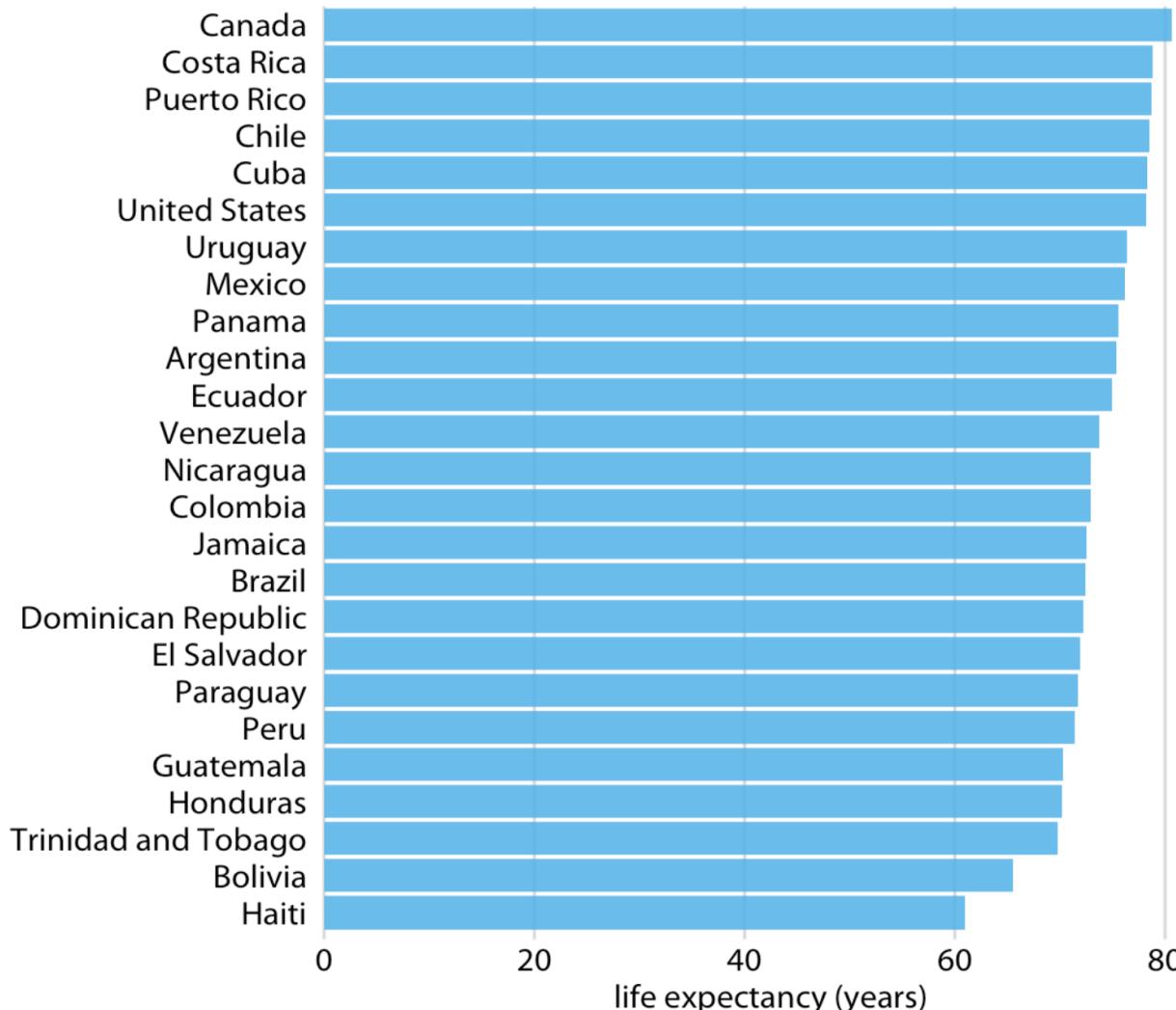
Barplots are assumed to start at zero

- Because bars are taken as physical objects
- Pros: the bar length \propto the amount shown
- Cons: maybe impractical or obscure some features, e.g. Uruguay, Mexico, Panama, and Argentina

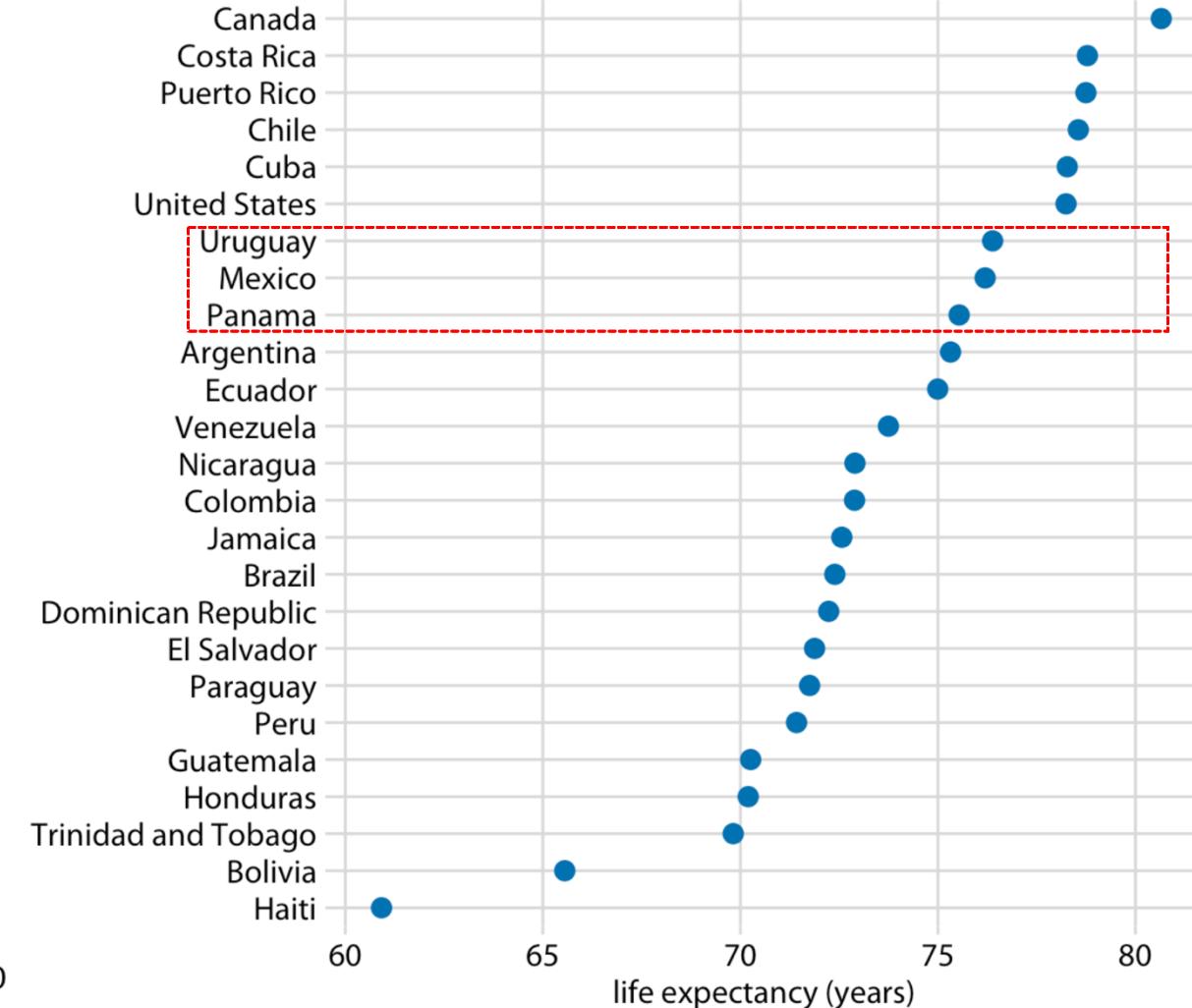


Barplots are assumed to start at zero

25 countries in the Americas

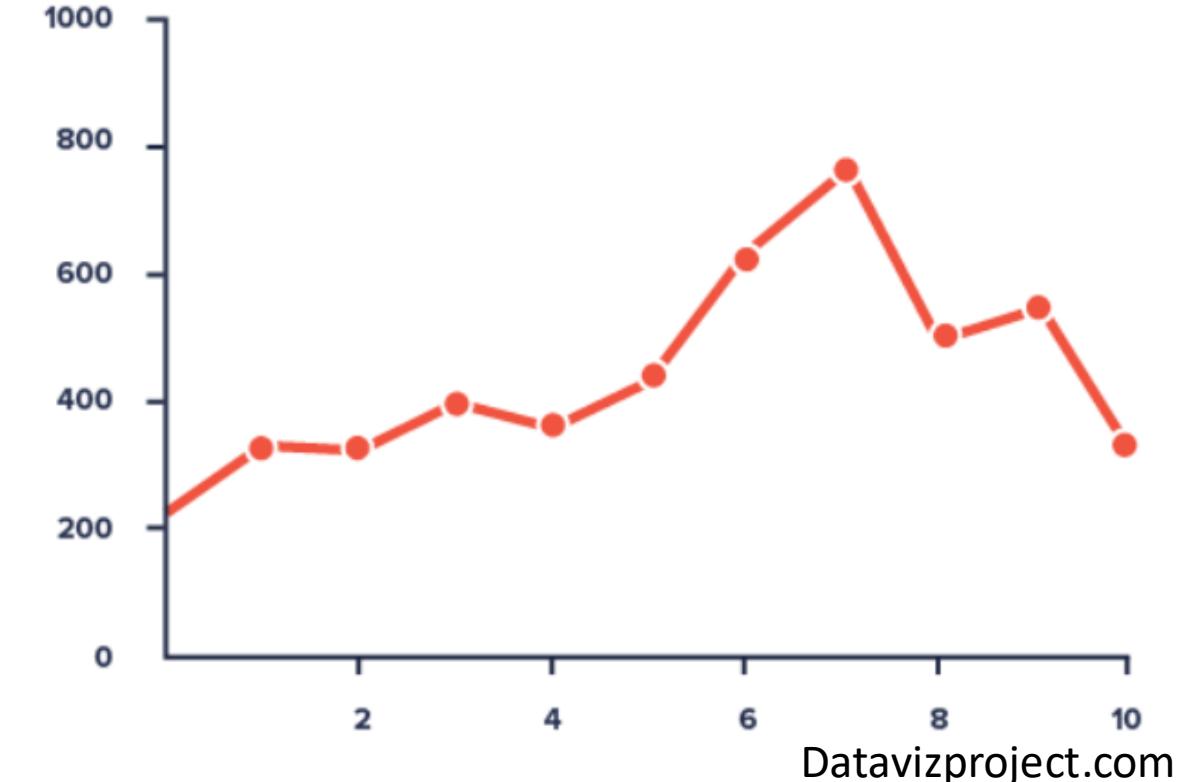
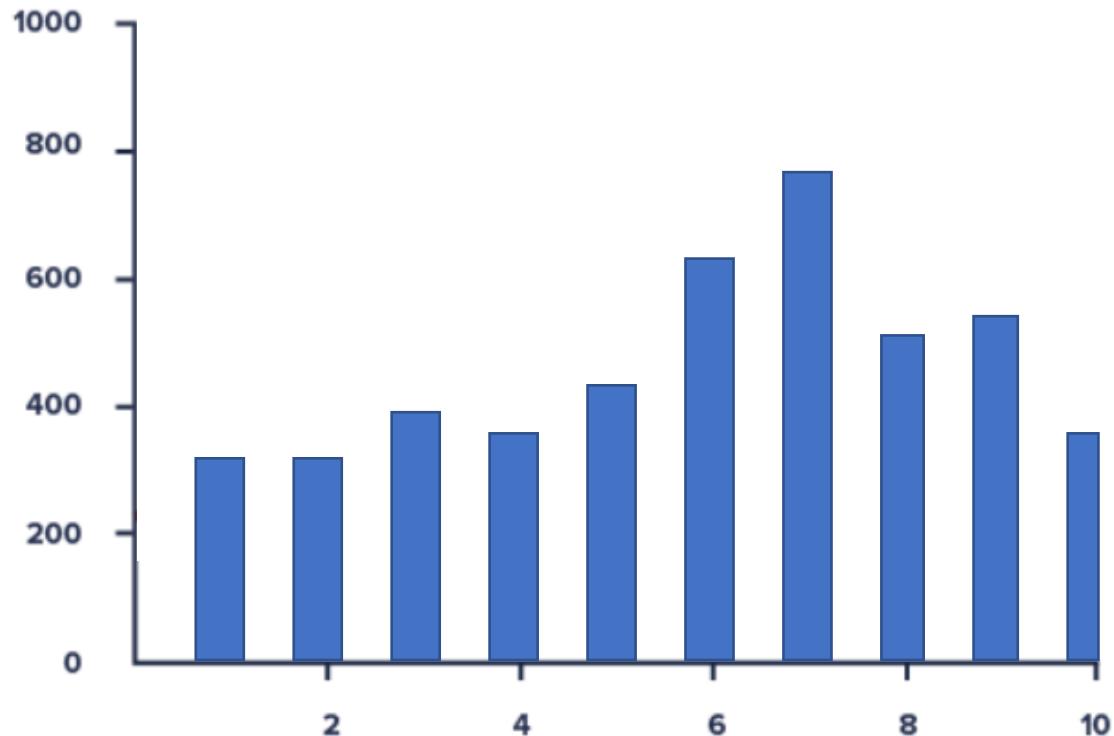


25 countries in the Americas



Line graphs show trends for amounts

- Barplots are for static values
- Line graph = dot plot (actual data) + real line (trend)
- To track changes along the x-axis (e. g., time)

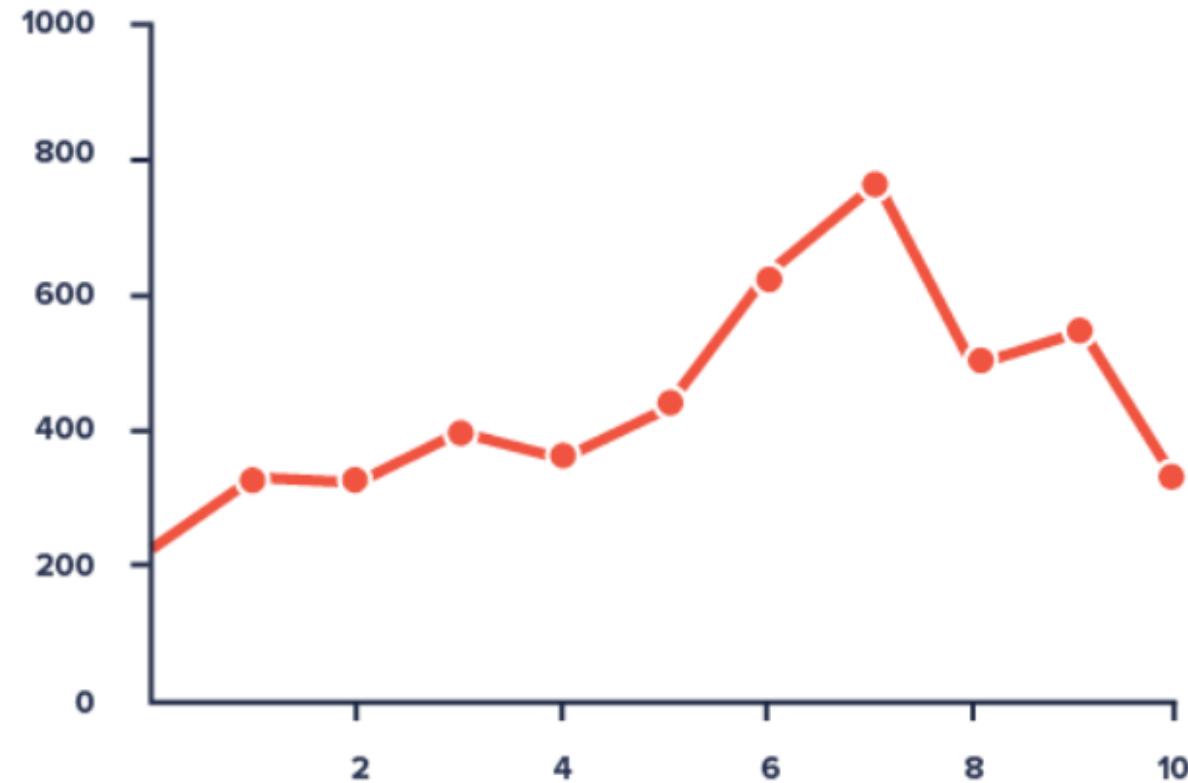
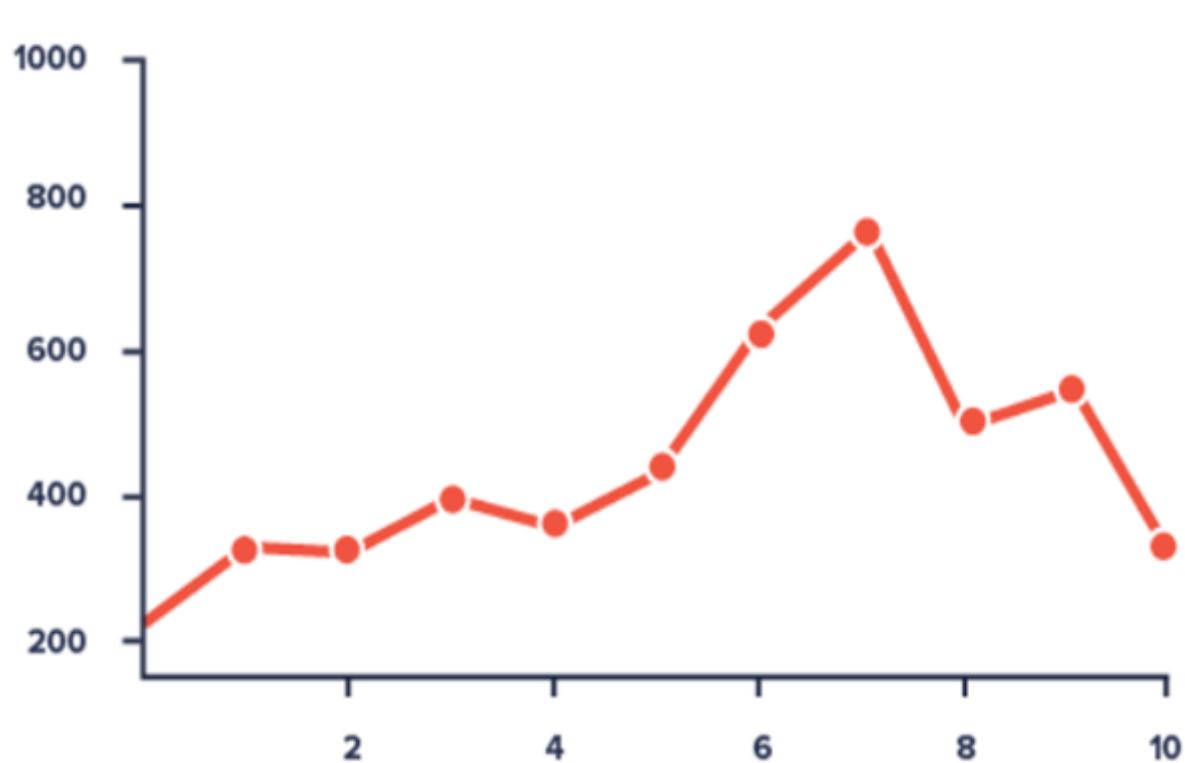


Datavizproject.com

Need to choose between barchart and line graph **with purpose**

Include a zero baseline if possible

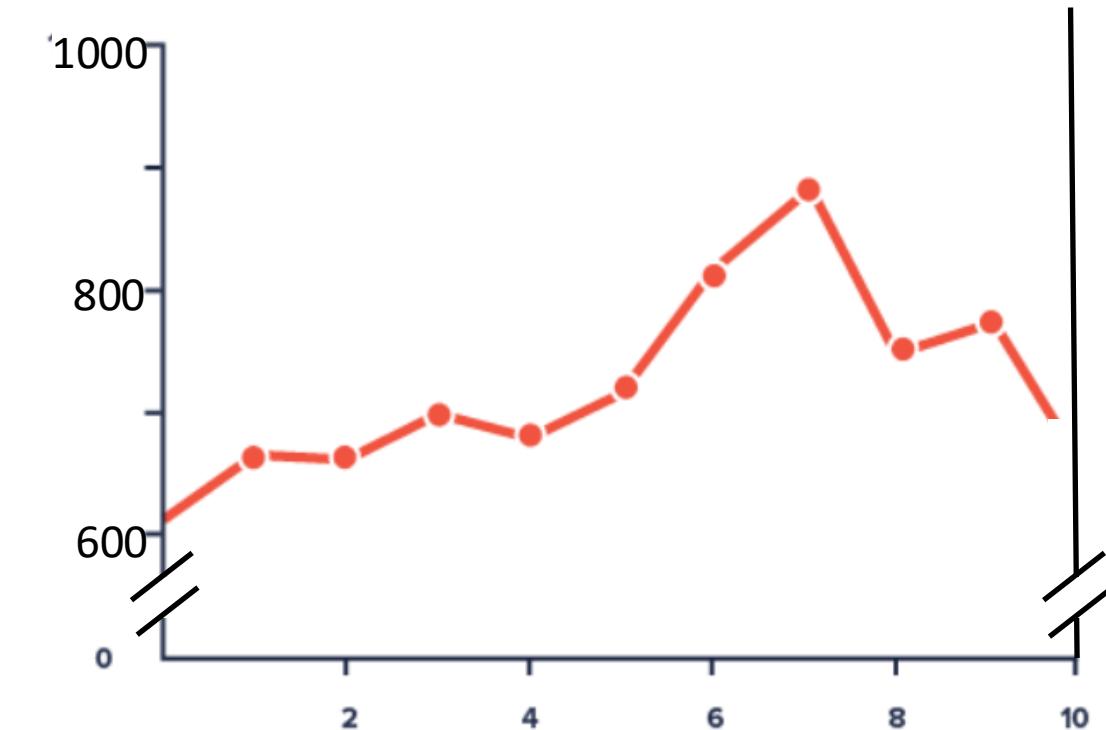
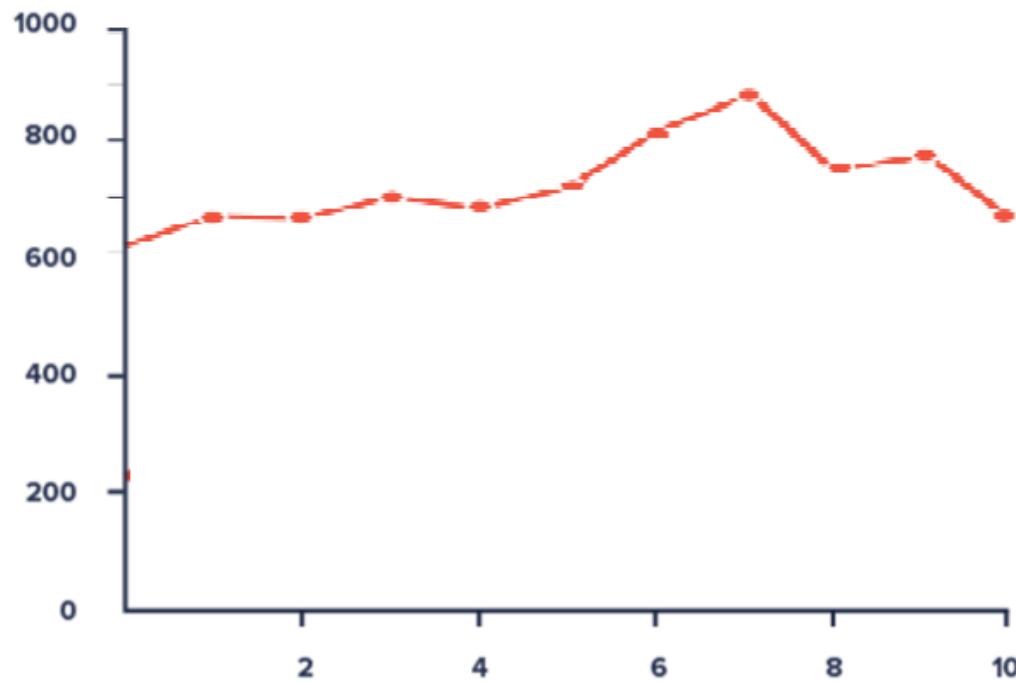
- A line graph does not strictly have to
- But better if it does (e.g., compare y-values for $x=3$ and $x=7$ in the two plots)



If not possible, truncate the axis

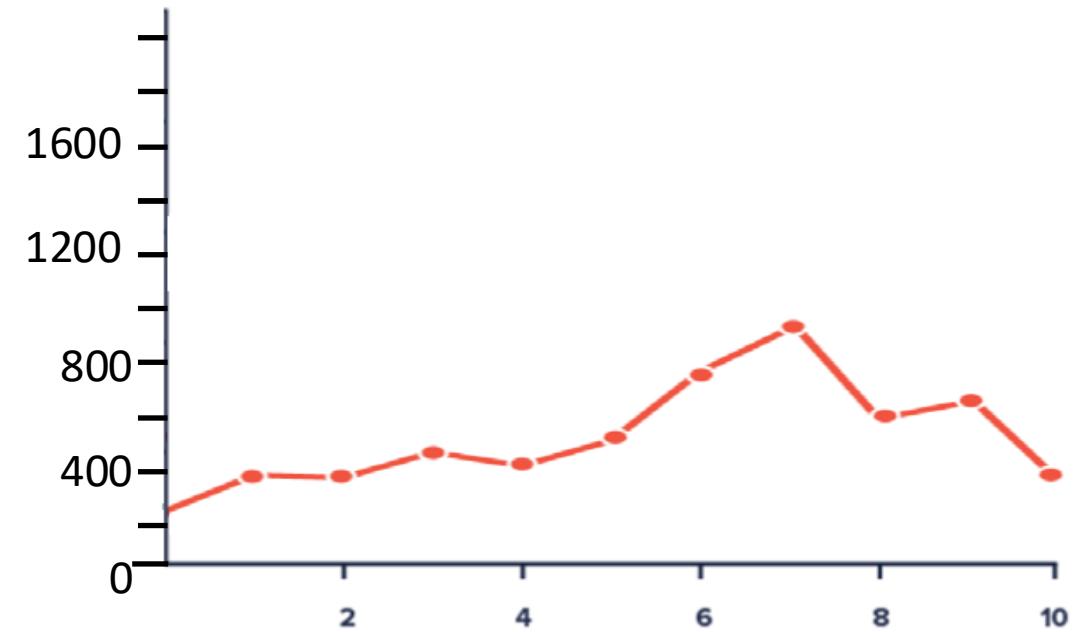
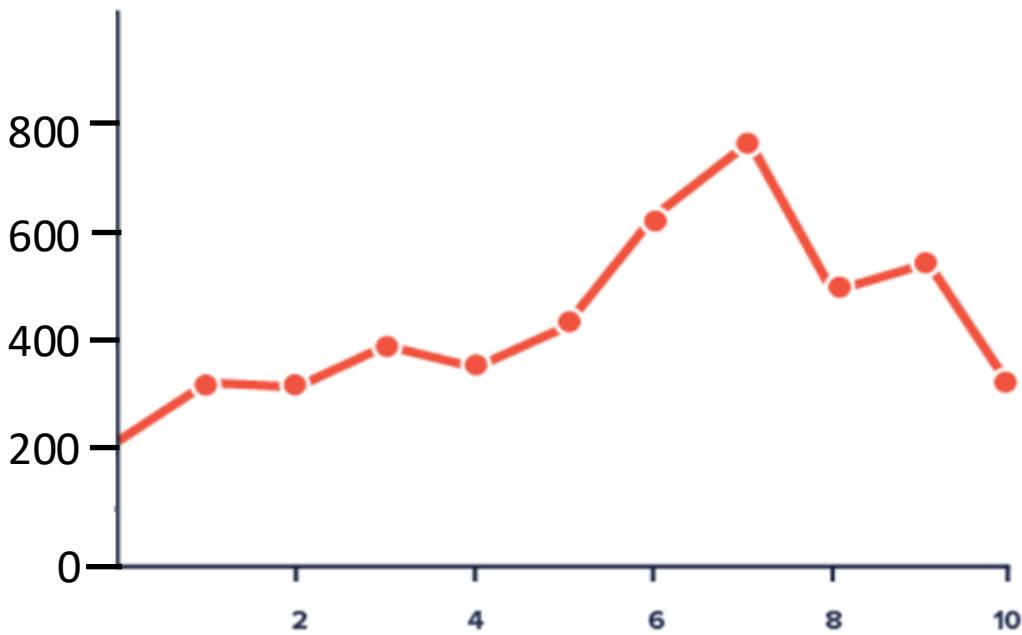
It is not possible when, e.g.,

- All data uniformly far from 0
- Small fluctuations matter



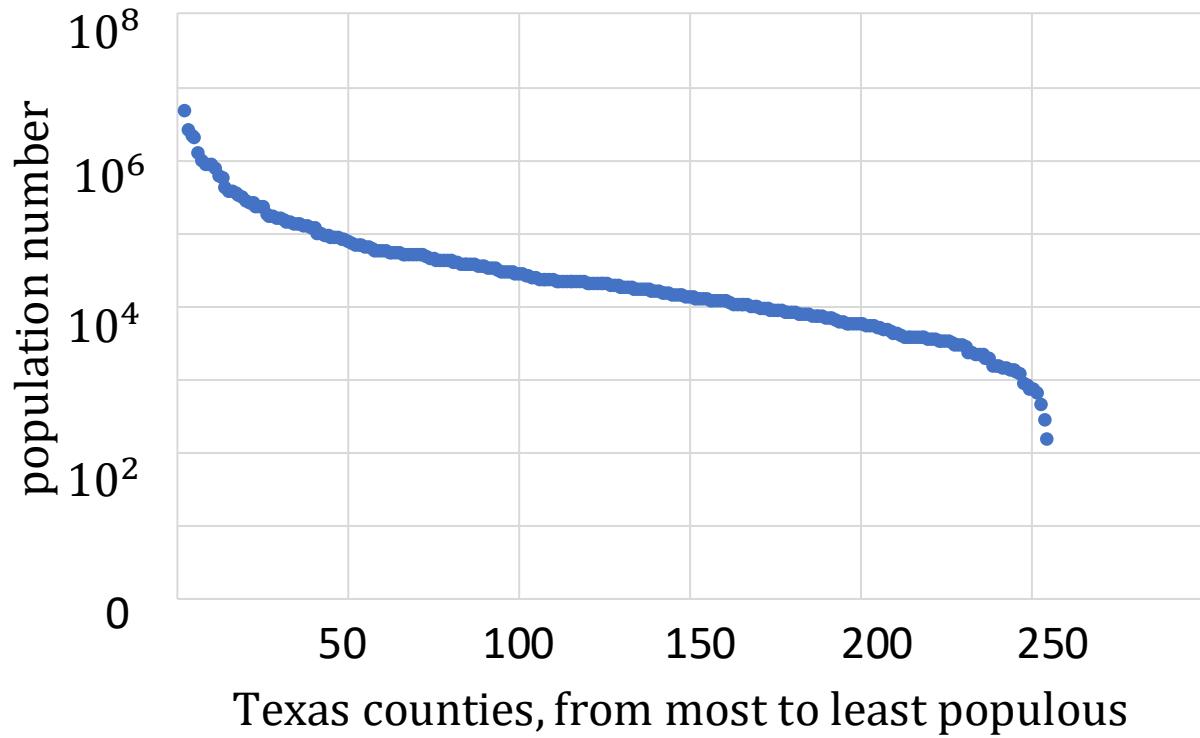
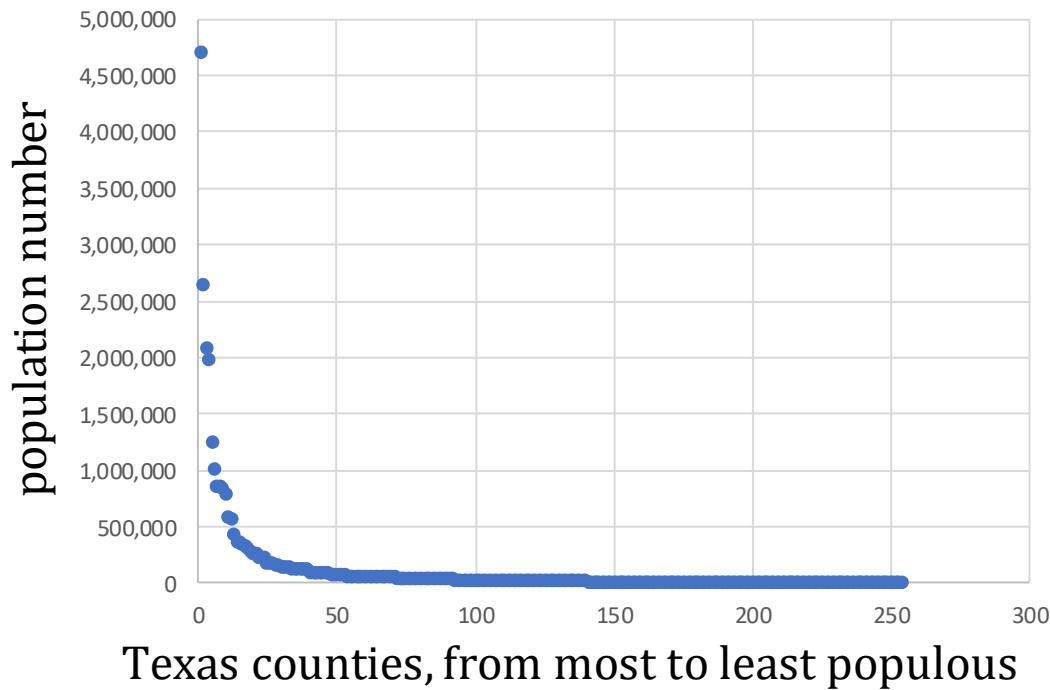
Use the right y-axis limit

- Ideally, the linegraph takes up 2/3 of the y-axis's total scale
- If not, you can modify the scale



Use the right height

- If not, you may want to modify the y-axis scale
- e.g., use a logarithm (base 10 here)



Bubble charts to visualize 3 dimensions

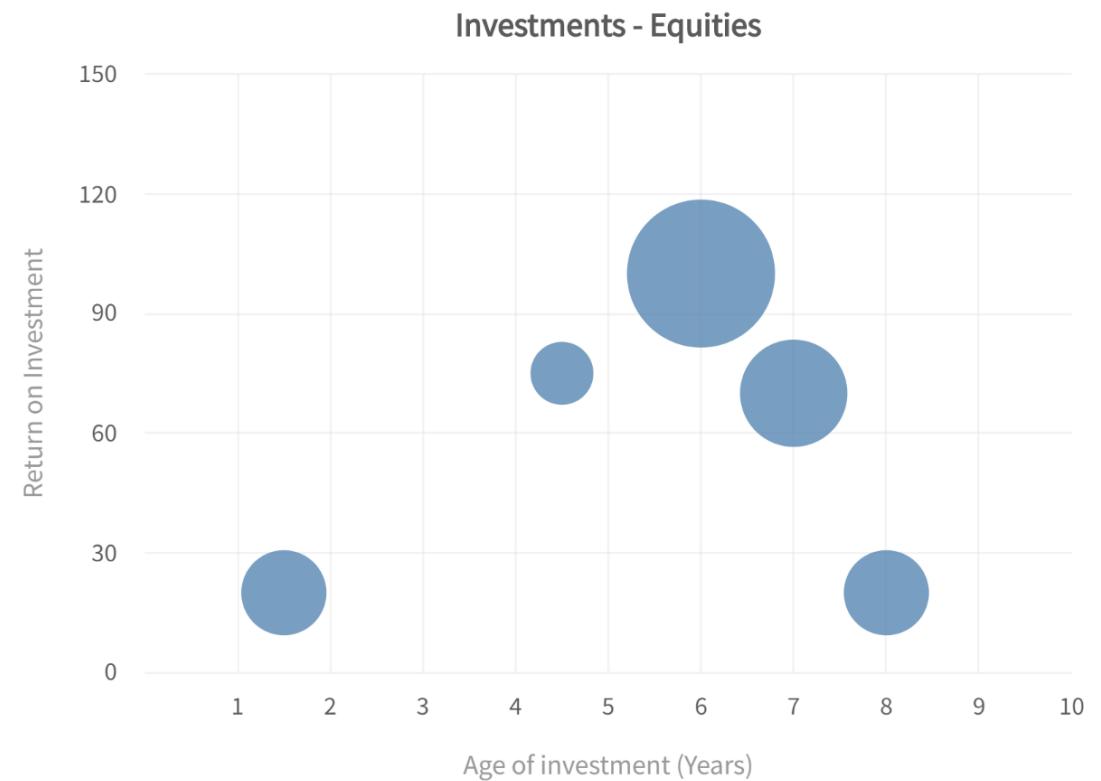
- Given three dimensions of data to visualize, how to visualize on a 2D plot?
- Put three column charts side by side? Why not?
- Use two dot plots? Why not?

Investments to 5 equities

	Age of investment (Year)	Return on investment	The amount of investment (1B\$)
A	1.5	22	2
B	4.5	75	1
C	6	100	4
D	7	70	3
E	8	21	2

Bubble charts to visualize amounts in 3 dimensions

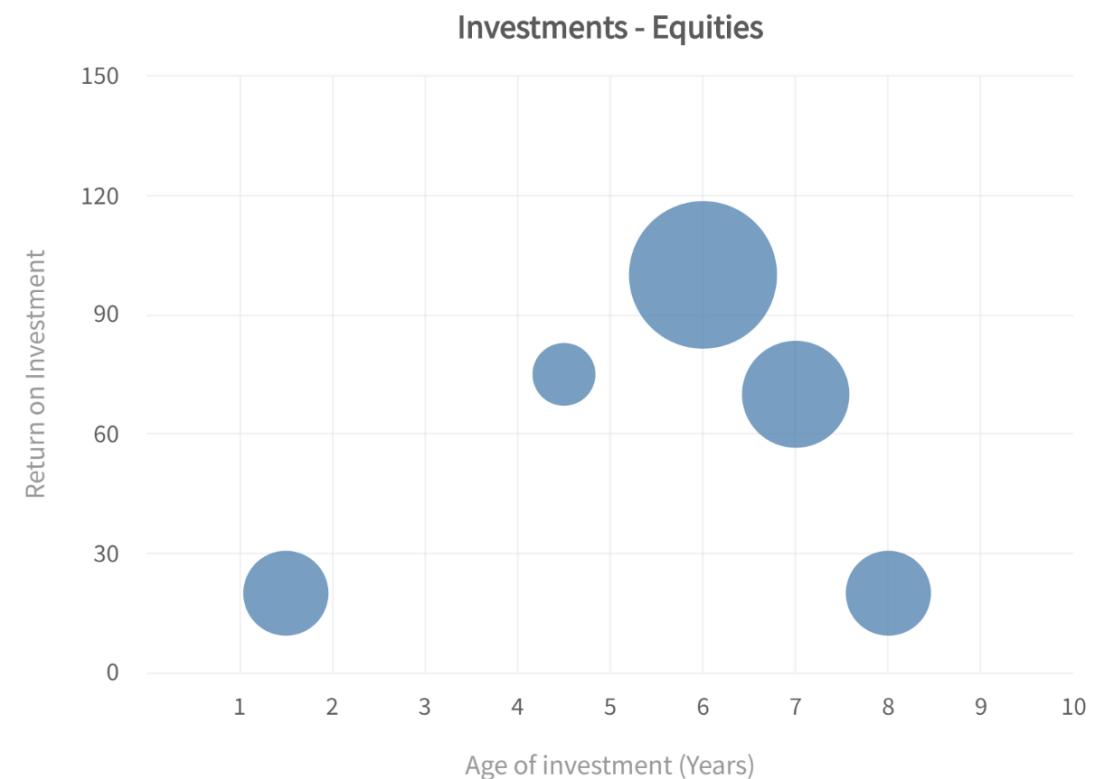
	Age of investment (Year)	Return on investment	The amount of investment (1B\$)
A	1.5	22	2
B	4.5	75	1
C	6	100	4
D	7	70	3
E	8	21	2



Bubble charts to visualize amounts in 3 dimensions

Bubblechart

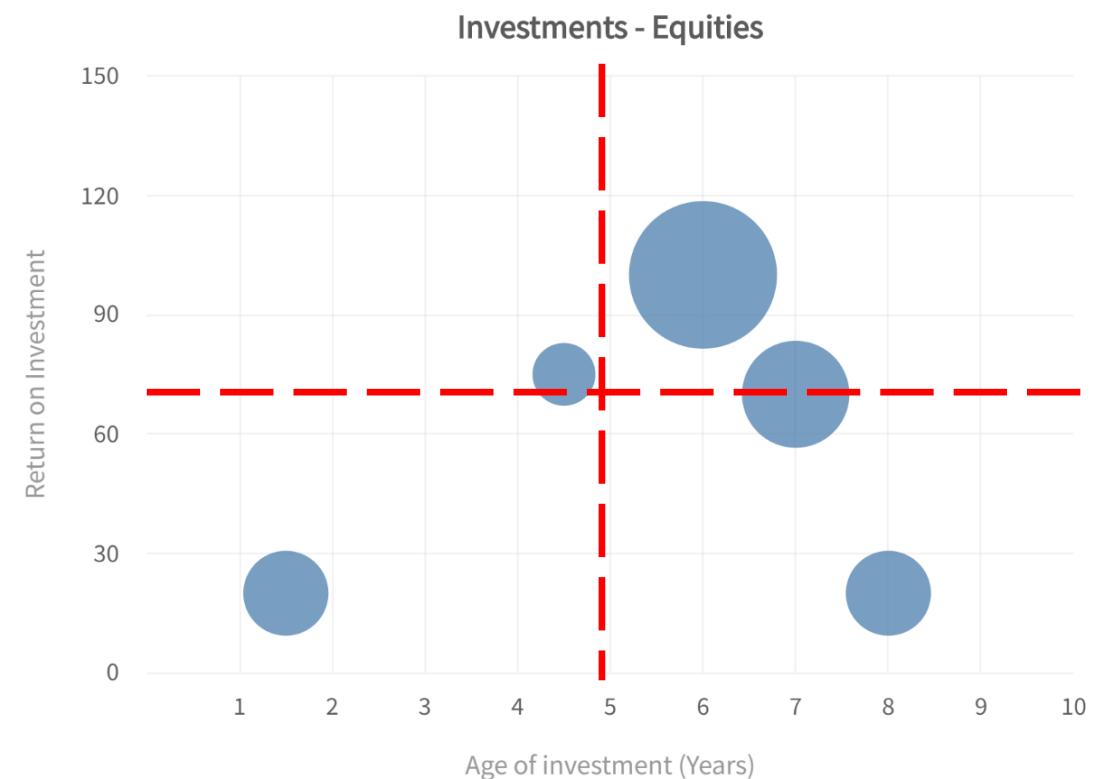
- To compare and depict the variable relationship by the position (1st and 2nd) and the size (3rd)



Bubble charts to visualize amounts in 3 dimensions

Bubblechart

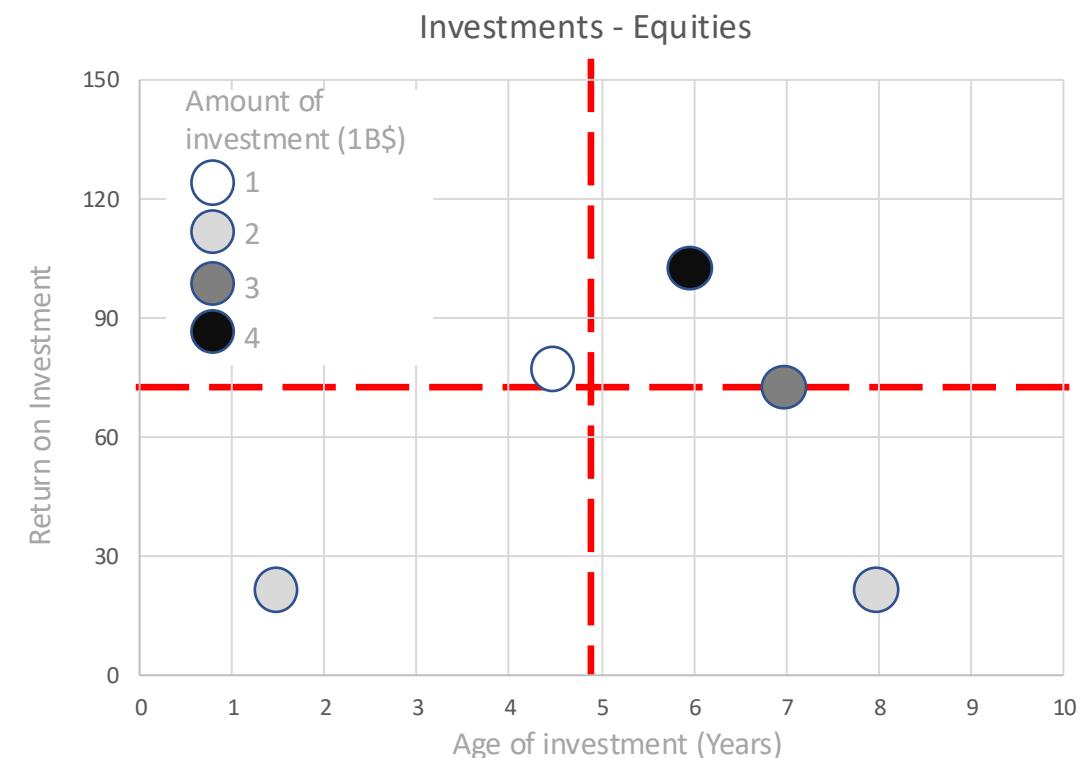
- To compare and depict variable relationships by the position (1st and 2nd) and the size (3rd)
- Also to analyze, e.g.,
 - Best investments
 - Good investments
 - Not-so good investments
 - Decent investments



Bubble charts to visualize amounts in 3 dimensions

Bubblechart

- To compare and depict variable relationships by the position (1st and 2nd) and the size (3rd)
- Also to analyze, e.g.
 - Best investments
 - Good investments
 - Not-so good investments
 - Decent investments
- The 3rd dimension also can be in bubble color

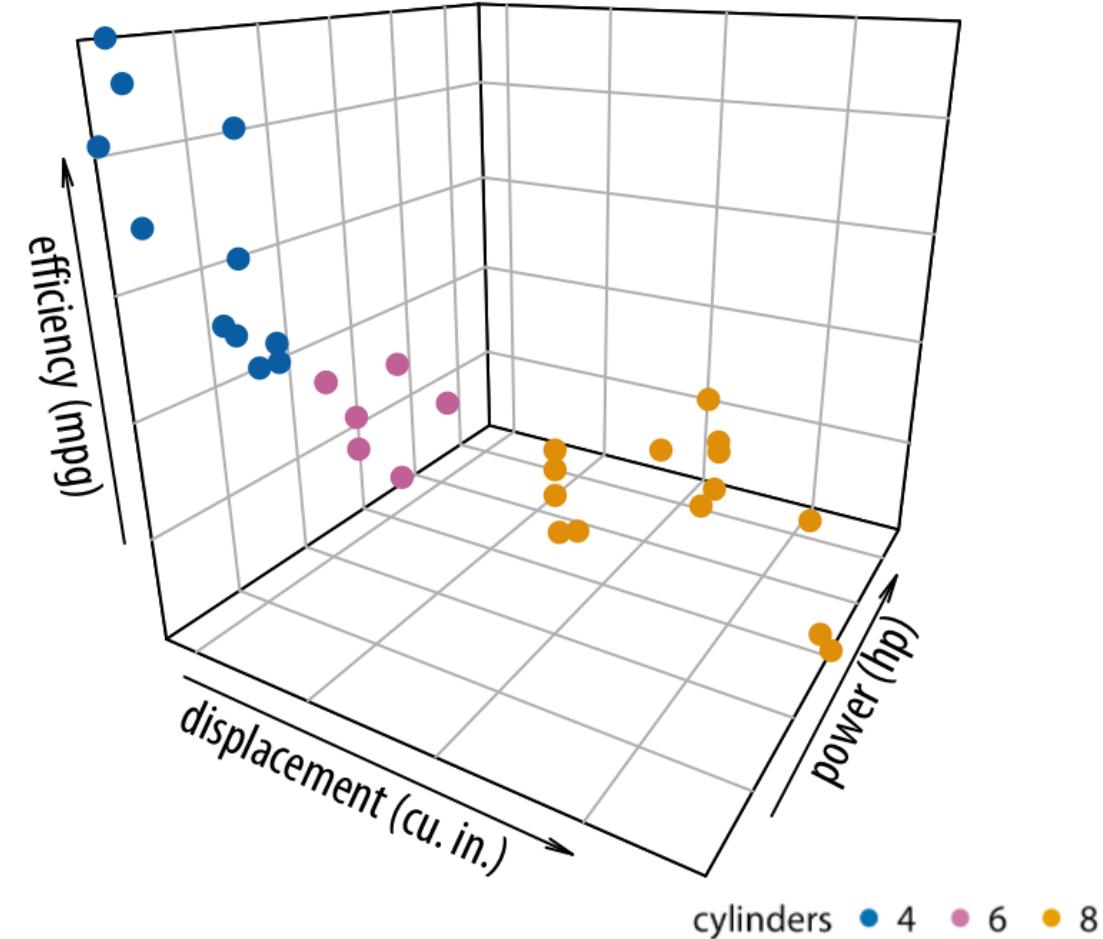


Bubbles chart avoid representation in 3D

Difficult to envision where the points are in space

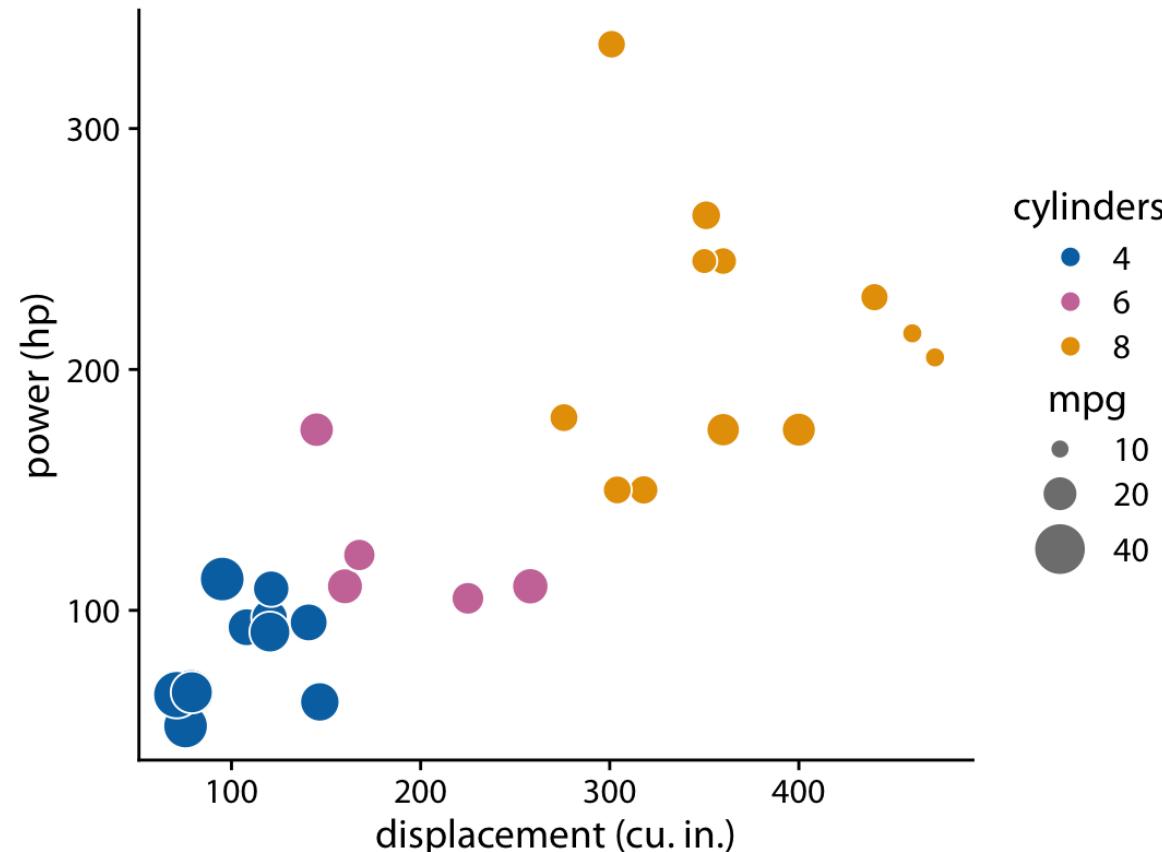
1. Data transformation from the data space into the 3D
2. Data transformation from the 3D to 2D problematic, since
 - non-invertible (a point in 2D ↔ a line in 3D)
 - Our attempt unreliable

Fuel efficiency vs. displacement and power for 32 cars



Bubble chart can represent 3D

Fuel efficiency vs. displacement and power for 32 cars.



For overlaps, you can

- Control transparency levels
- Give jittering

Bubble size scaled in area

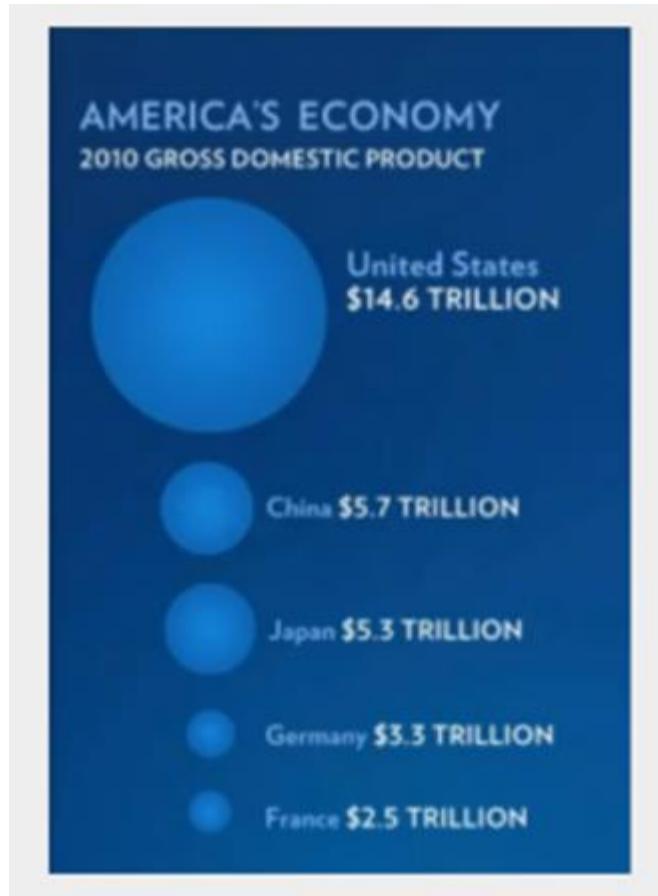
- In the example, the radius of circle scaled with the size of economy
- The radius scales linearly, but the area scales quadratically



The 2011 State of the Union Address

Bubble size should be in area

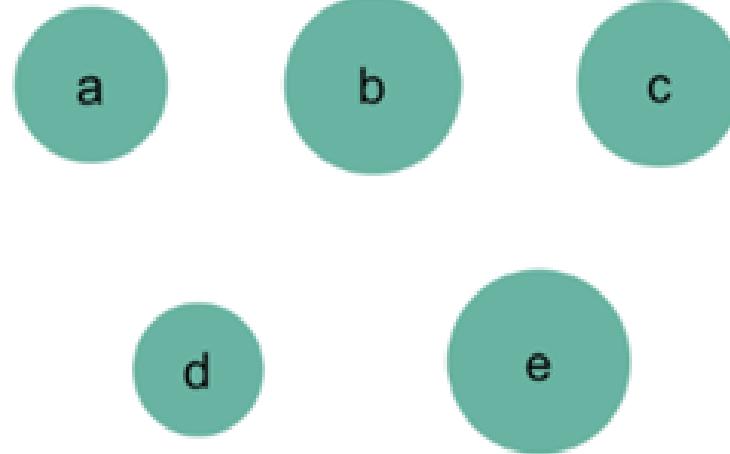
Before and after correction



Bubble sizes difficult to compare

e.g. Data points of size 17, 24, 20, 15, 27

In bubble area size



In barplot



- Used with a scaling factor
- Attention usually goes to the biggest and the smallest

Lollipop to start at 0 while highlighting endpoints

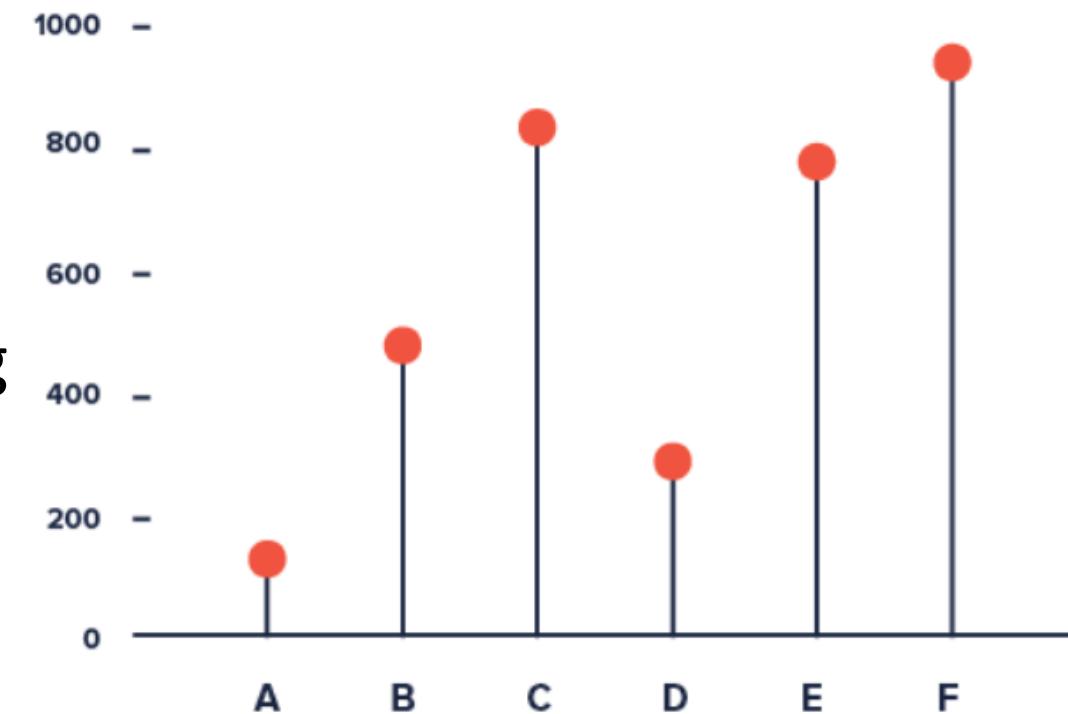
Lollipop (=a line and a dot) combines the advantages of

- Box: direct representation easily perceived
- Dot: flexible for radical changes in a wider range

e.g. a large set of tall columns not intimidating

but with inherent inaccuracies due to

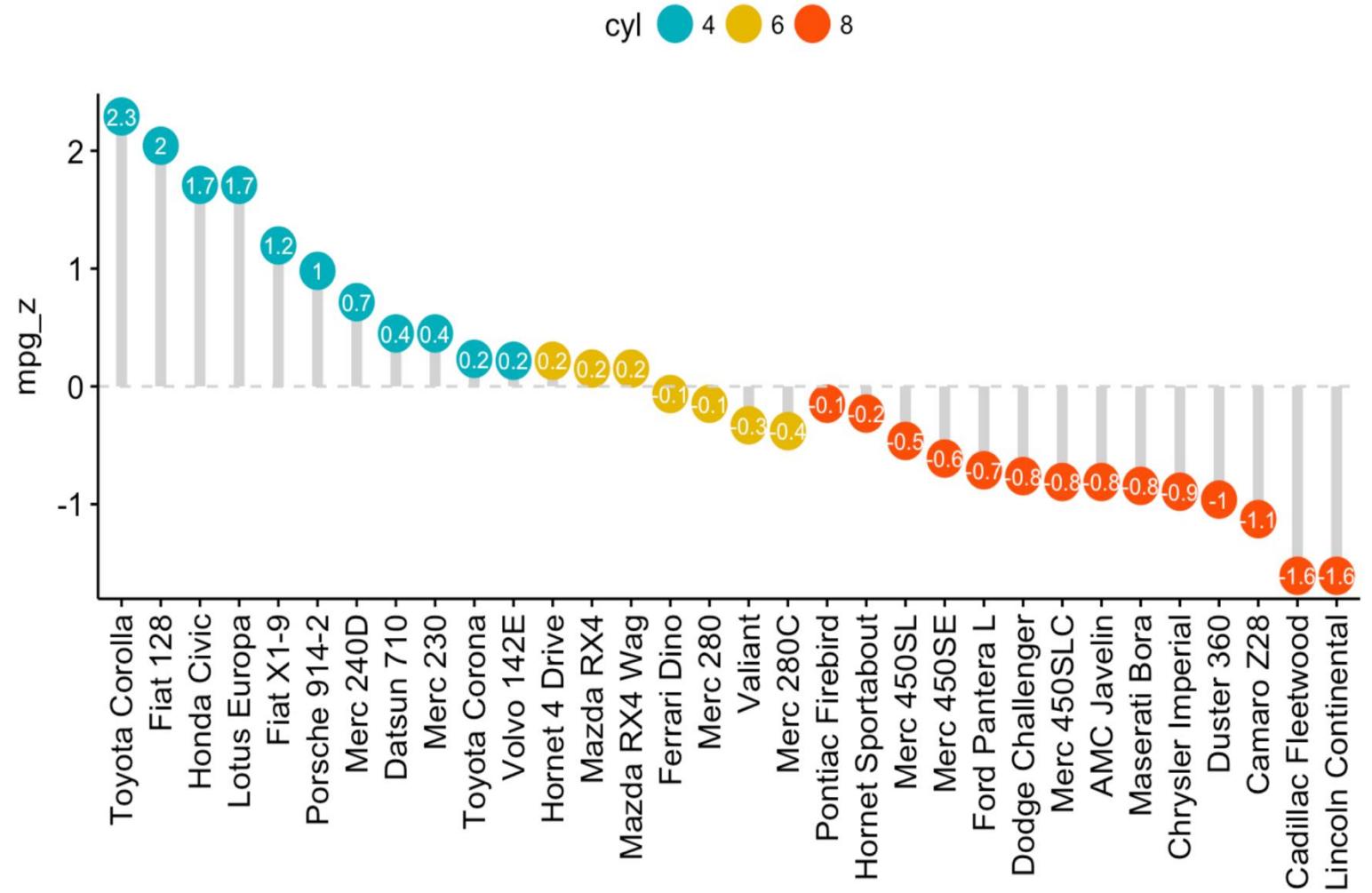
- The location of the center
- Half of the circle extending



Lollipop to start at 0 while highlighting endpoints

Lollipop (a line and a dot)
without the inherent
inaccuracies due to

- Circle center imprecision
- Important to sort

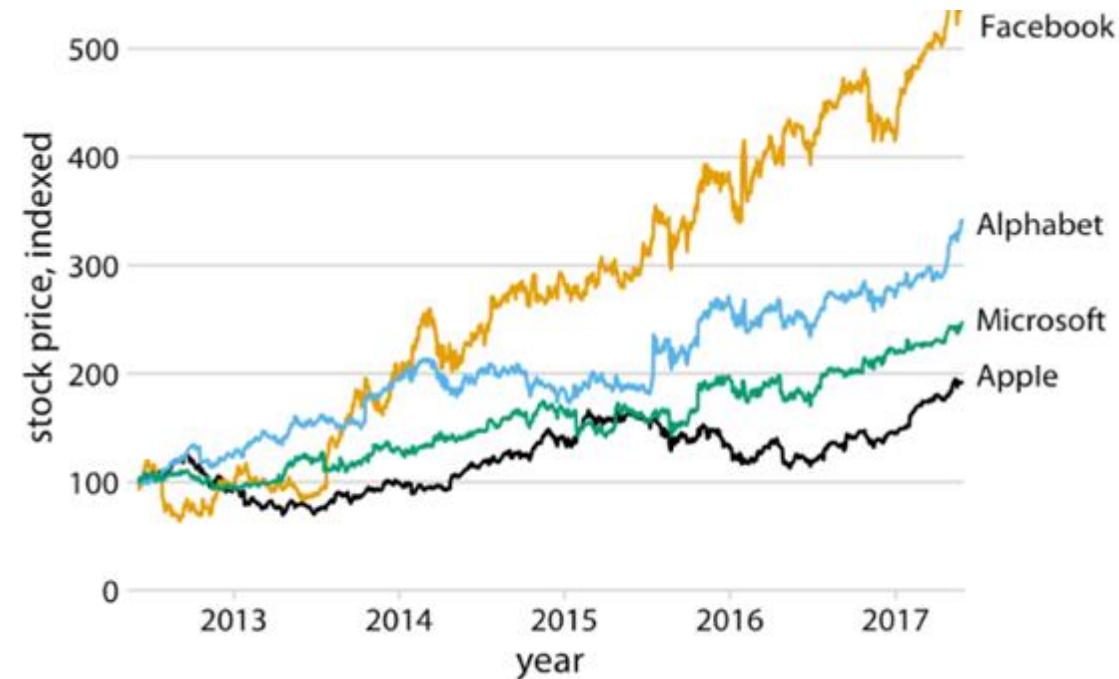
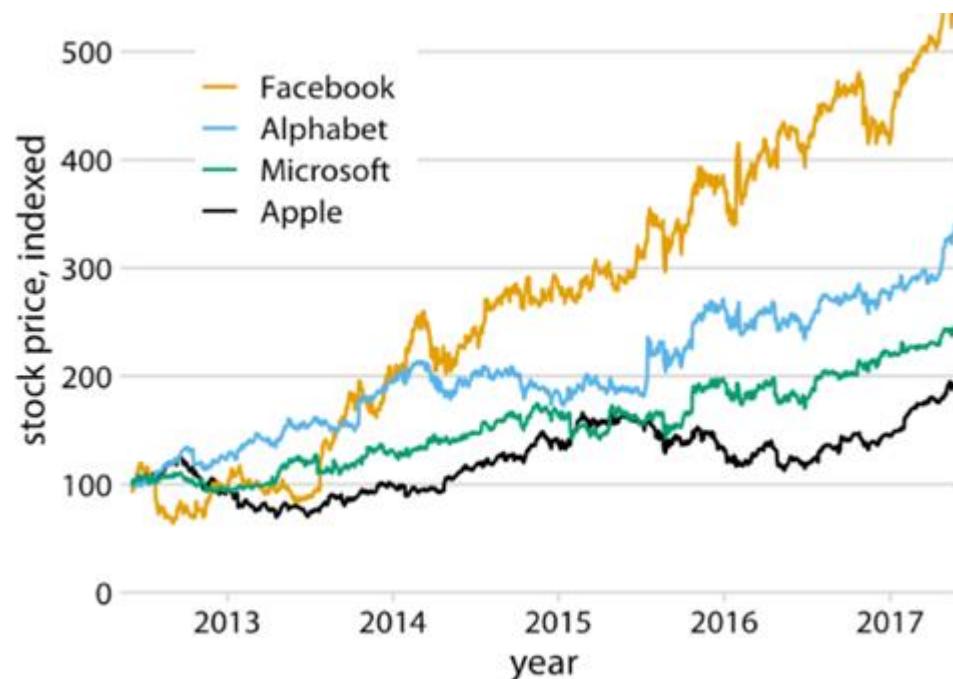


Adapted from <http://www.sthda.com/english/articles/24-ggpubr-publication-ready-plots>

Label lines directly

To avoid

- Referencing a legend (legend locations can vary)
- Obstructing signals



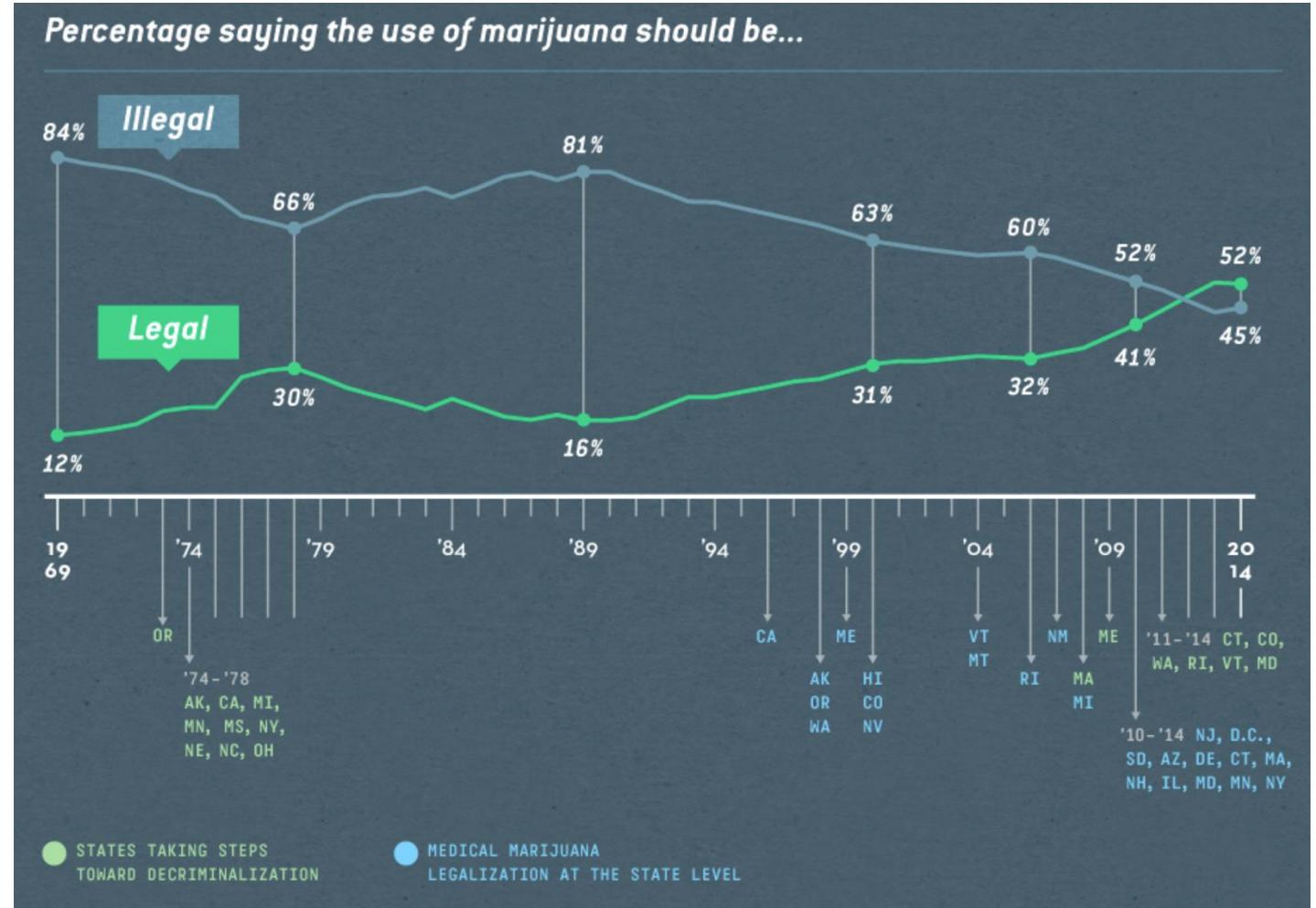
Distinguish points from lines in line graphs

Stylish lines can obstruct data points.

Which are real values?

Are all the datapoints indicated with dots?

Which values on the lines are just interpolations?



http://wp_media-prod.s3.amazonaws.com/app/uploads/2015/09/24113017/Infographic_Design_Example_Digit.png

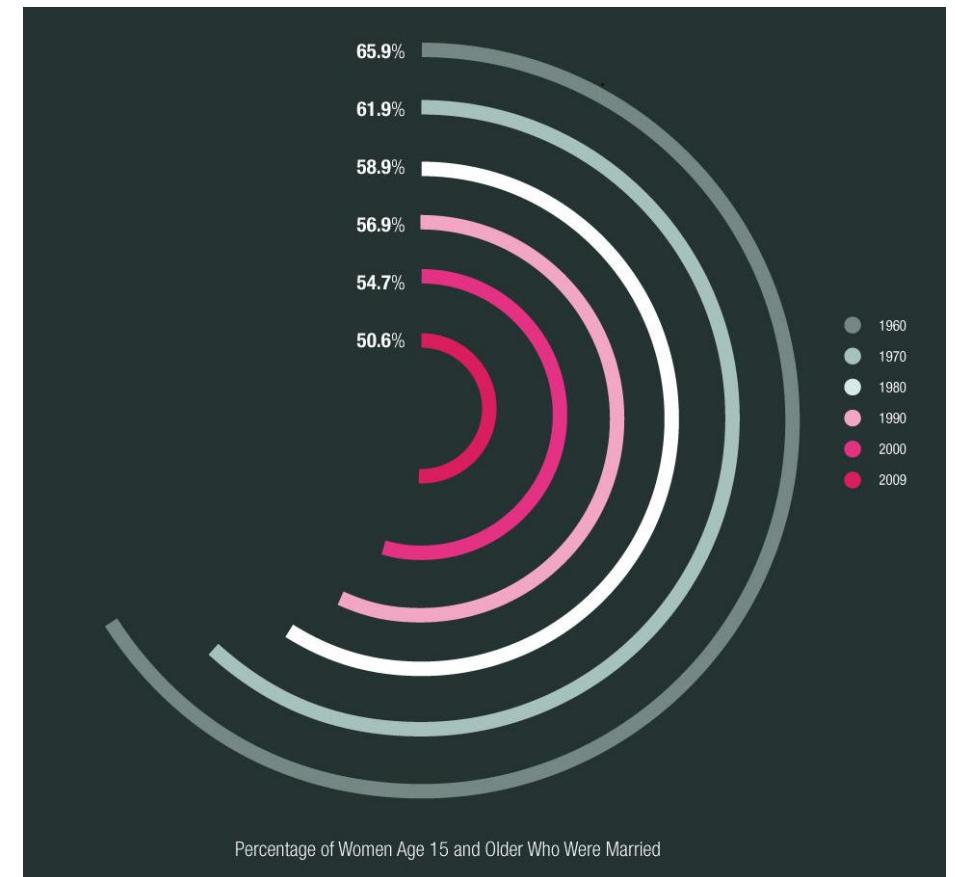
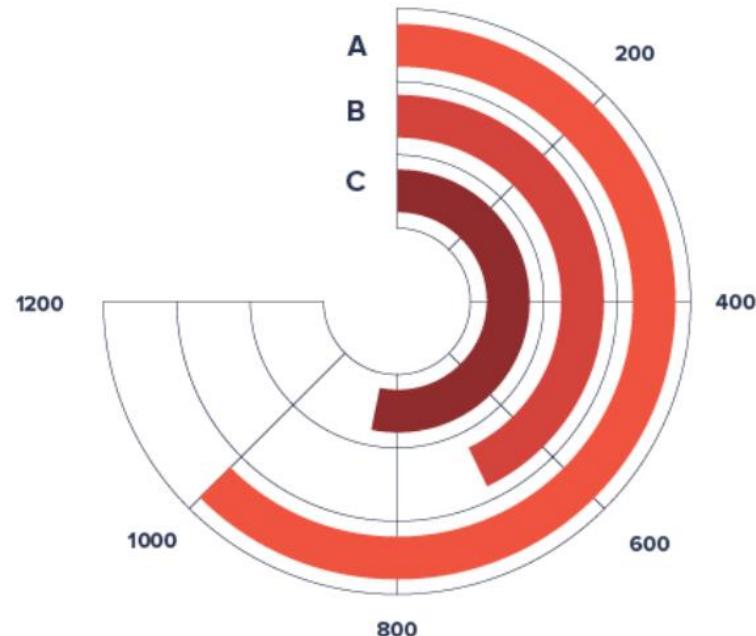
Radial Barchart

Barchart on radial axis

Pros: ?

Cons: deceptive, probably more difficult to read

Radial Bar Chart



Radial barchart limitations

- The bars are on arcs of different radii!
- So their lengths can't be compared directly!
Rescaling needed.
- https://www.data-to-viz.com/caveat/circular_barplot_accordion.html

