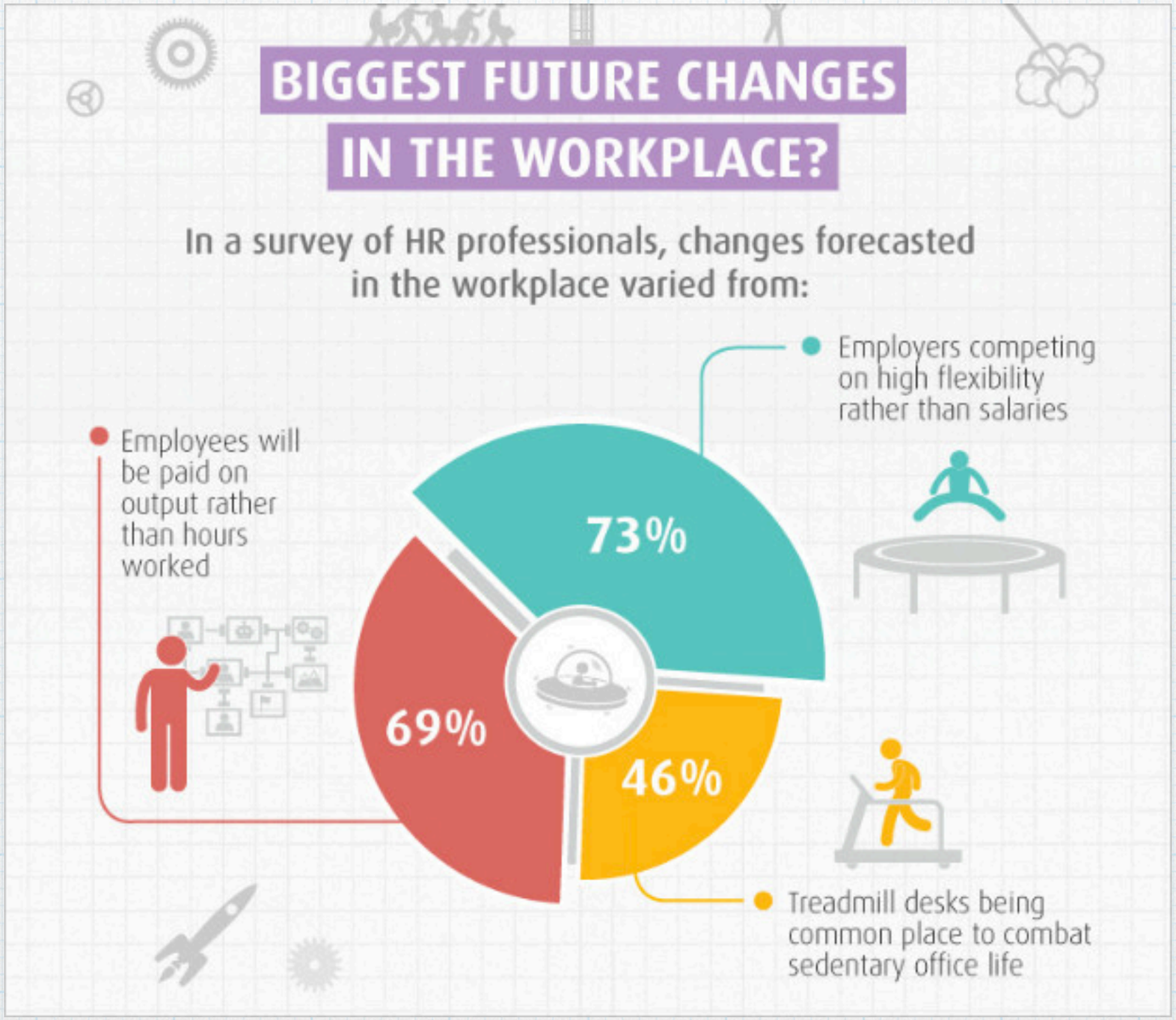
Visualization

CAC 350

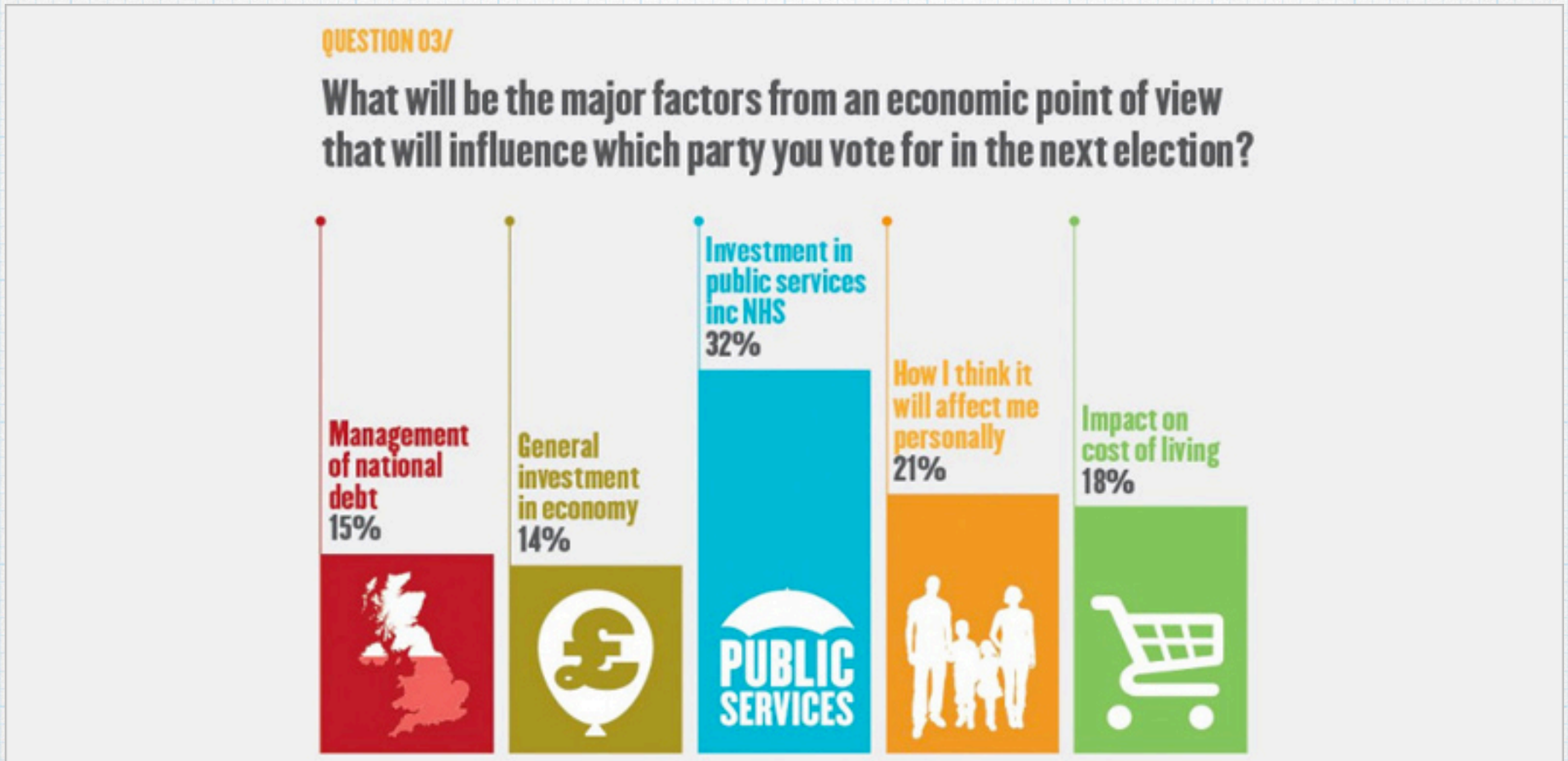
Effective Charts

* Convey the right and necessary information without distorting facts
* Simple in design, viewers shouldn't have to strain to understand the chart
* Aesthetics are supportive and not distracting
* Shouldn't contain too much information

Hmmm...pie chart. 100%?

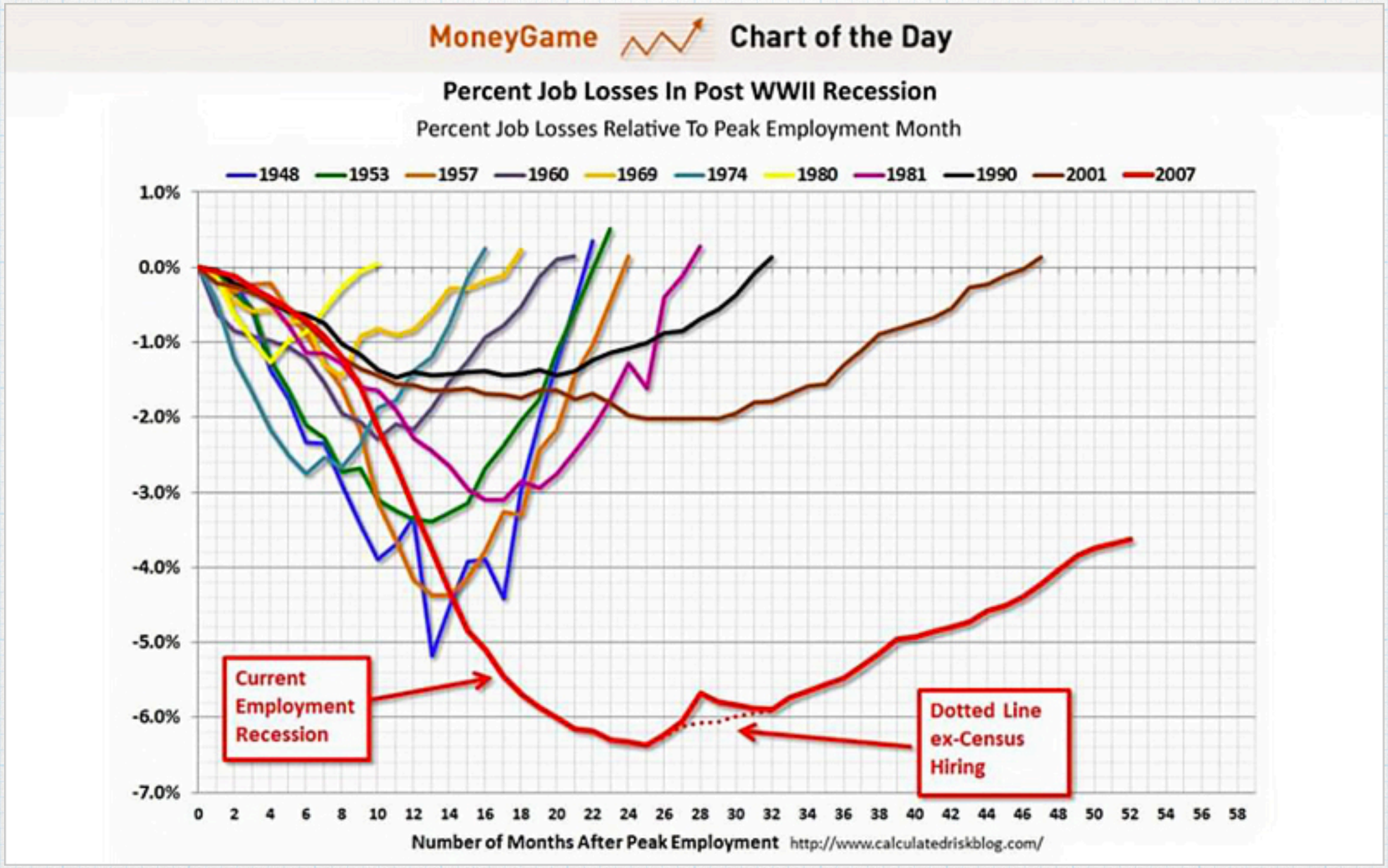


https://visme.co/blog/bad-infographics/



https://visme.co/blog/bad-infographics/

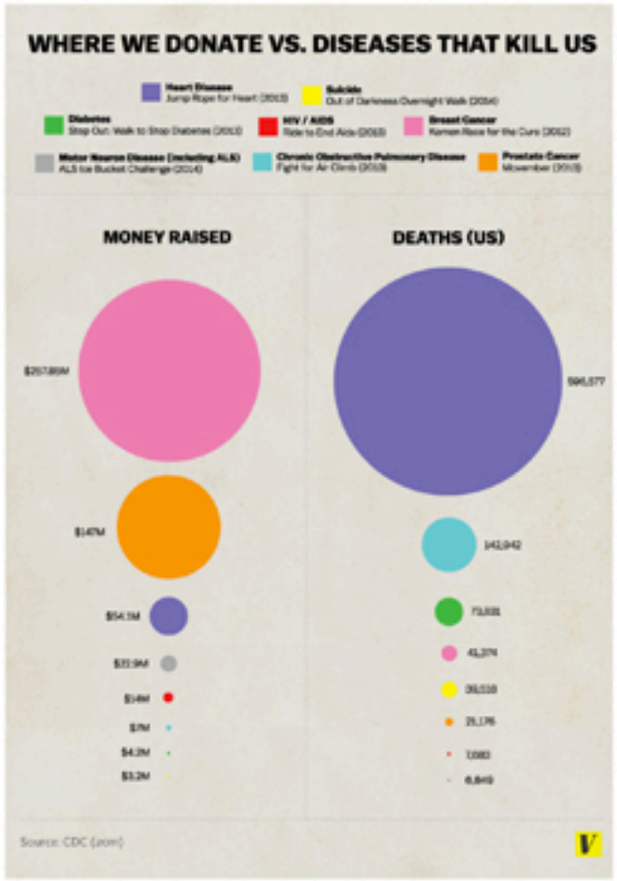
Percentage of a whole...pie chart or stacked bar would be more appropriate



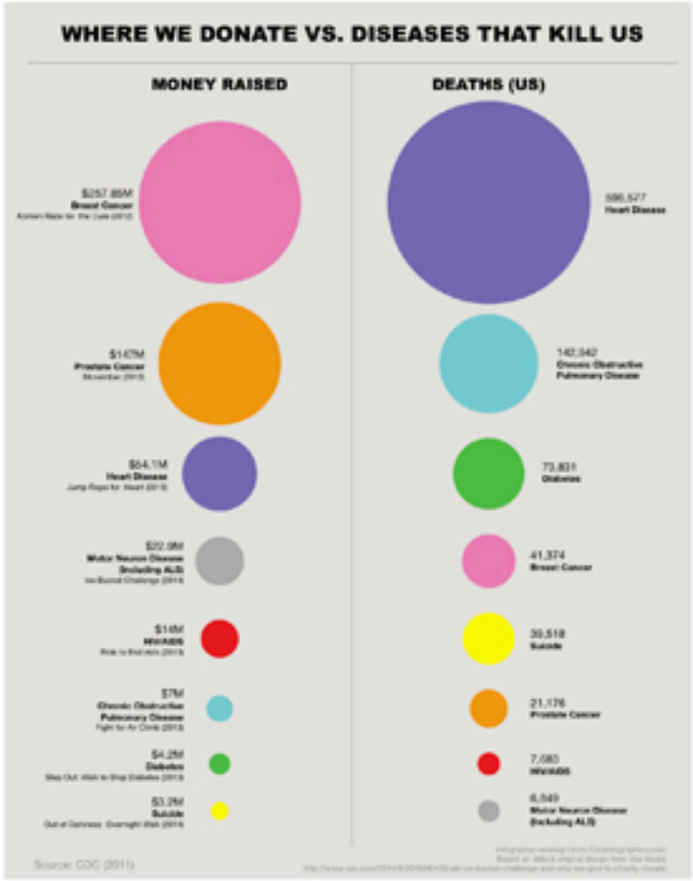
https://visme.co/blog/bad-infographics/

Too much information

Original Design



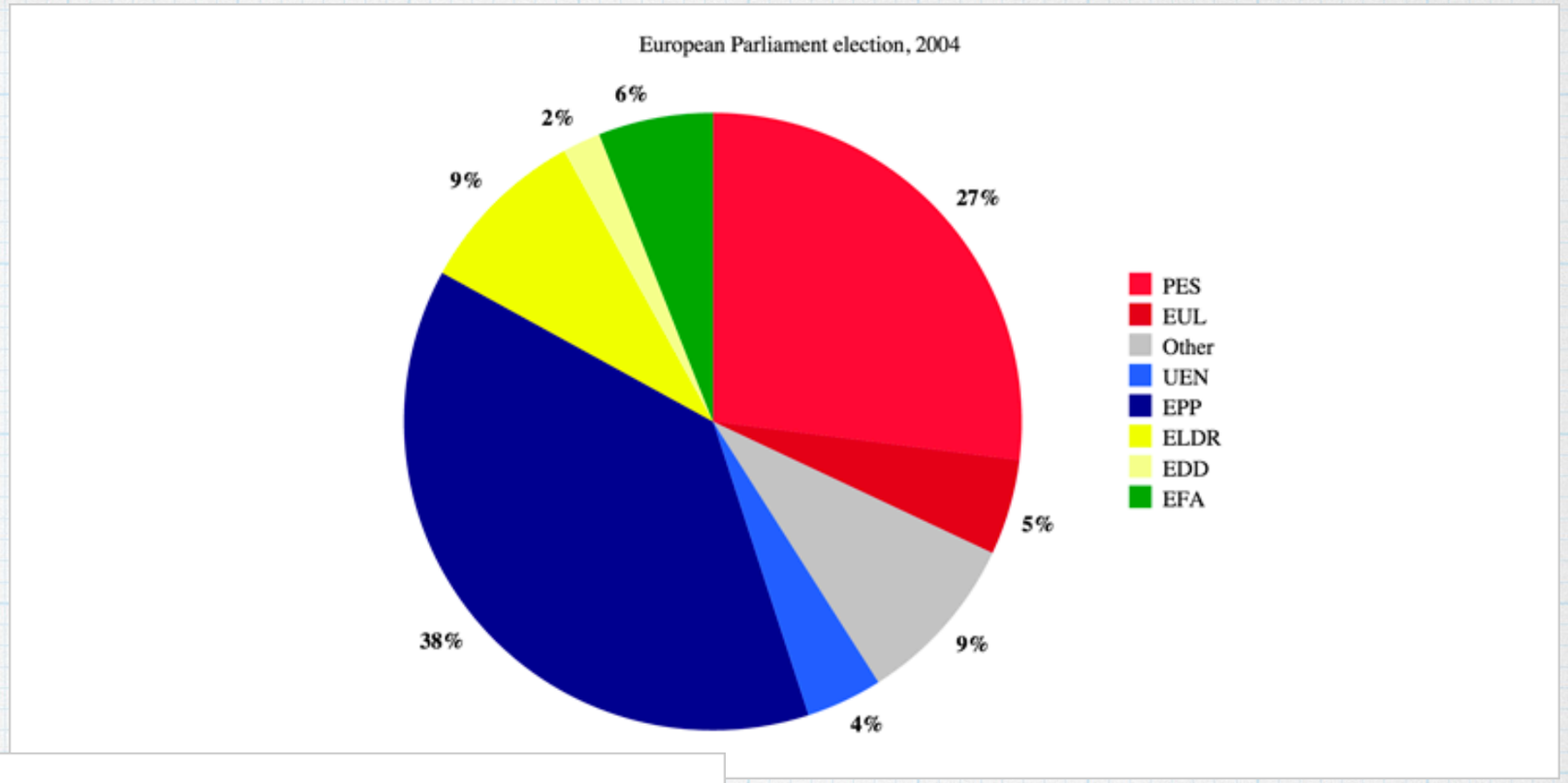
Corrected Design

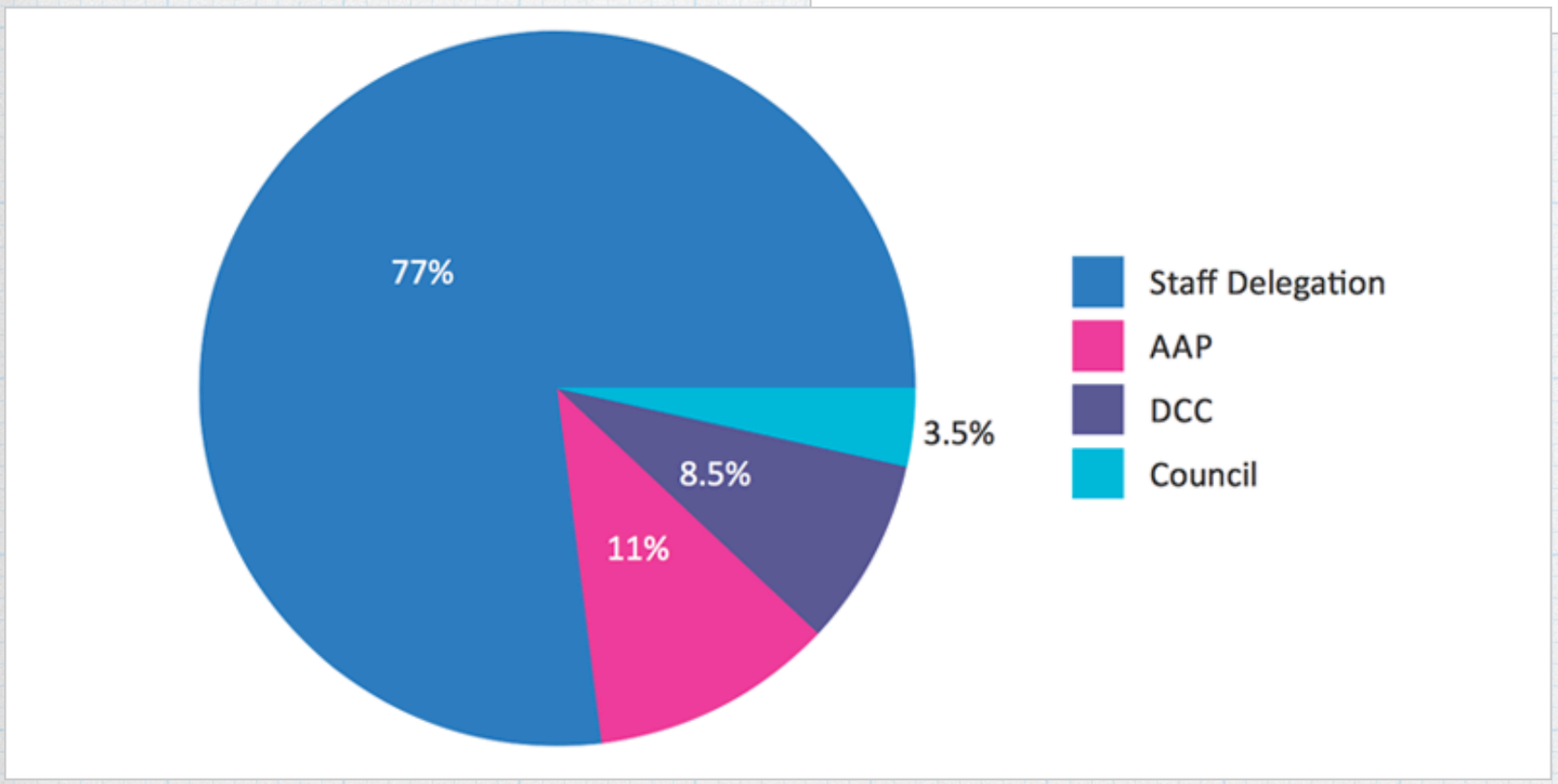


https://visme.co/blog/bad-infographics/

Inaccurate scales

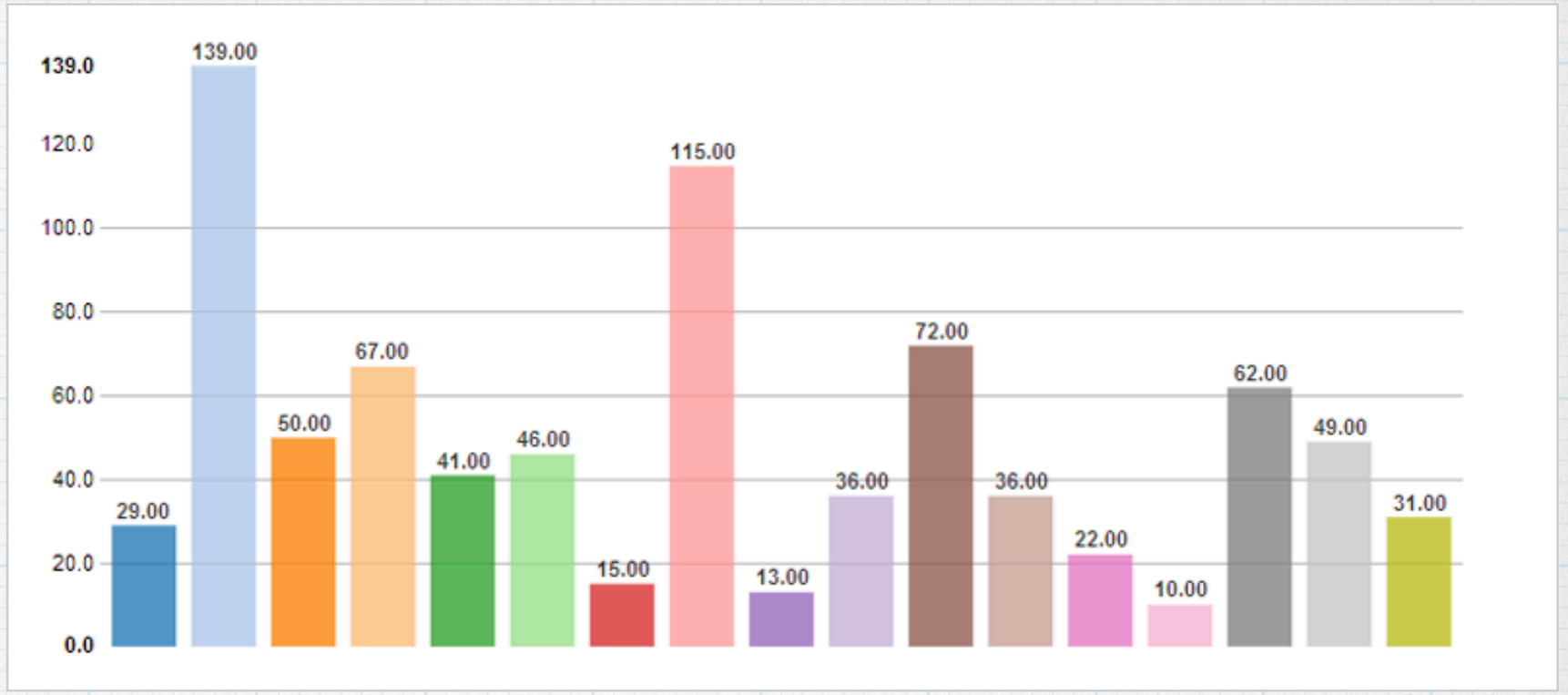
Organization matters

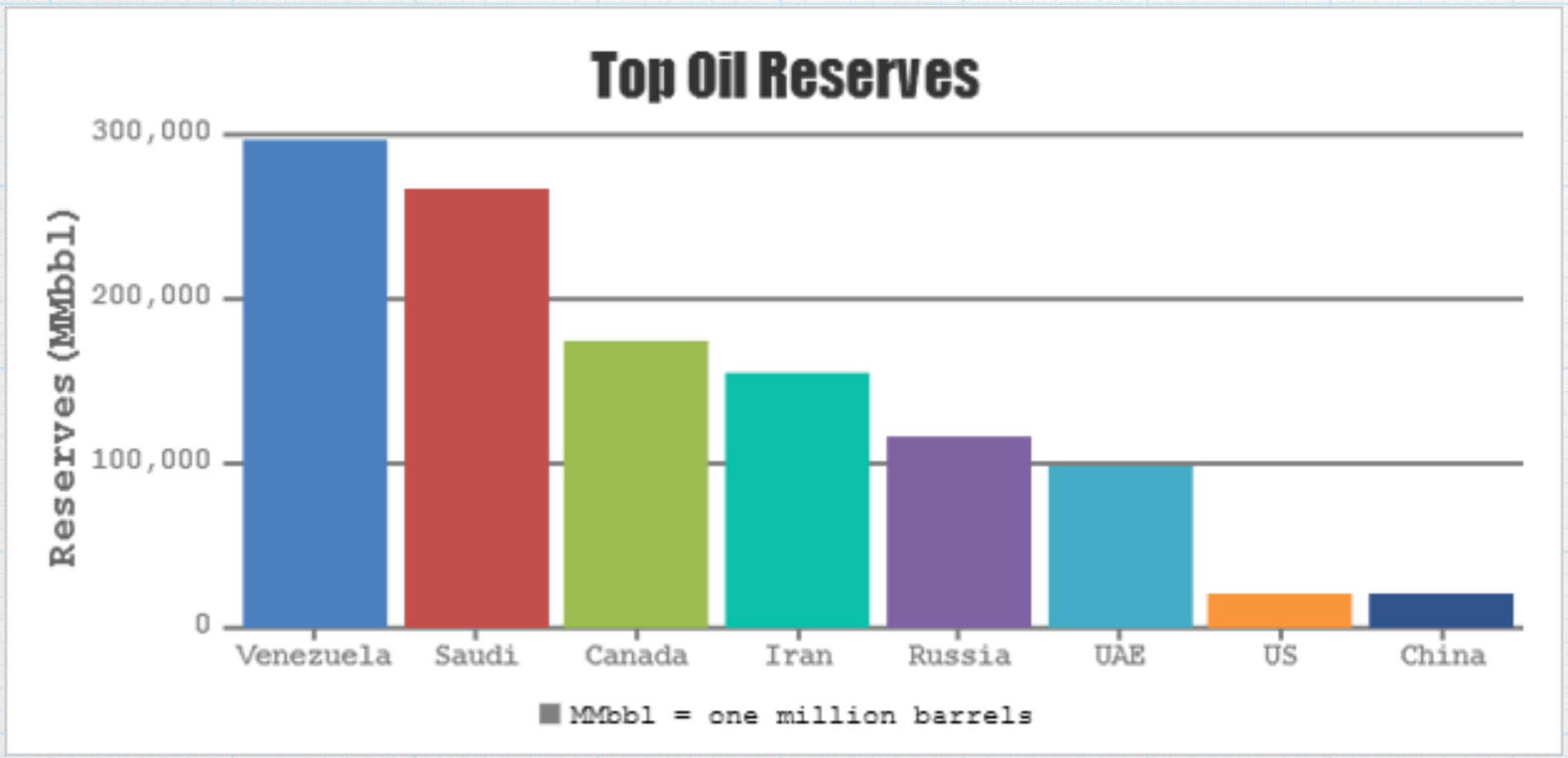




https://visme.co/blog/bad-infographics/

Organization matters





https://visme.co/blog/bad-infographics/

Correlation

* In statistics, correlation refers to the degree to which two random variables are related.

* Related plots: scatter plot, scatter plot with line of best fit, bubble plot, correlogram, marginal boxplot, pairwise plot

Scatter Plot

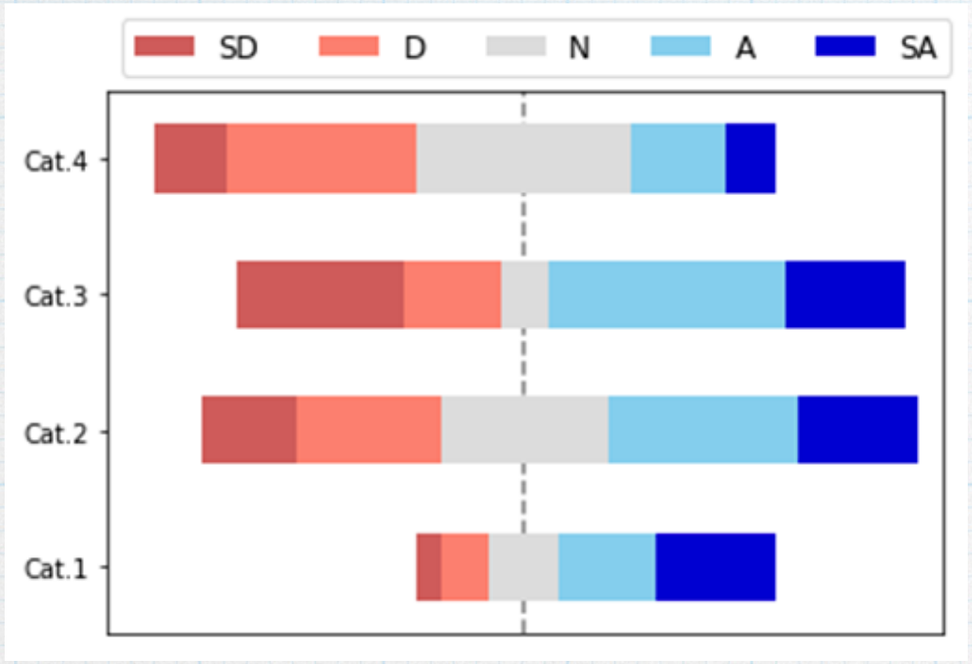
* plt.plot and plt.scatter (bubble plot)
* For larger datasets, plt.plot is faster (doesn't have to render each dot individually)

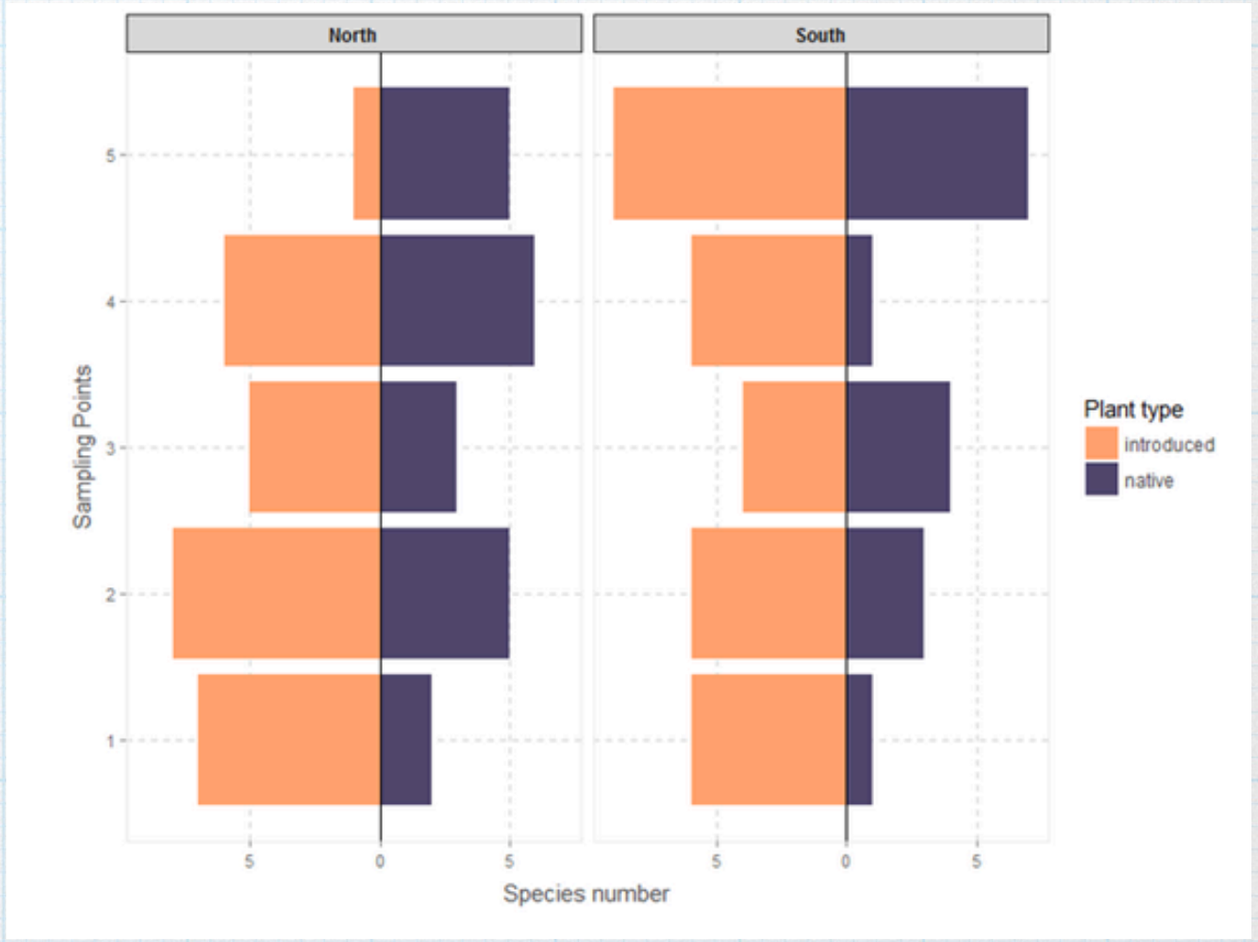
https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.scatter.html

Deviation

* Measure of the difference between an observed variable and some other value (usually the variable's mean)
* Common plots: diverging bars, diverging dot, area chart, box plot

Diverging Bar





Plotly is your best bet
Workaround presented

Distribution

* How does the data spread across the results
* Common plots: histogram, density plot, joy plot, box plot, violin plot

Your Turn

* Grab a partner and experiment
  * Grab the dataset called cars1 (on Moodle) and plot mpg, cylinders, and weight to illustrate correlation using your preferred chart
* Simulate data...you'll need 5 questions, likert scale (1-5) for each question, randomize responses (15 responses per question). Create a plot showing the deviation.
* Using the earthquake dataset, creatively display the distribution of the magnitude