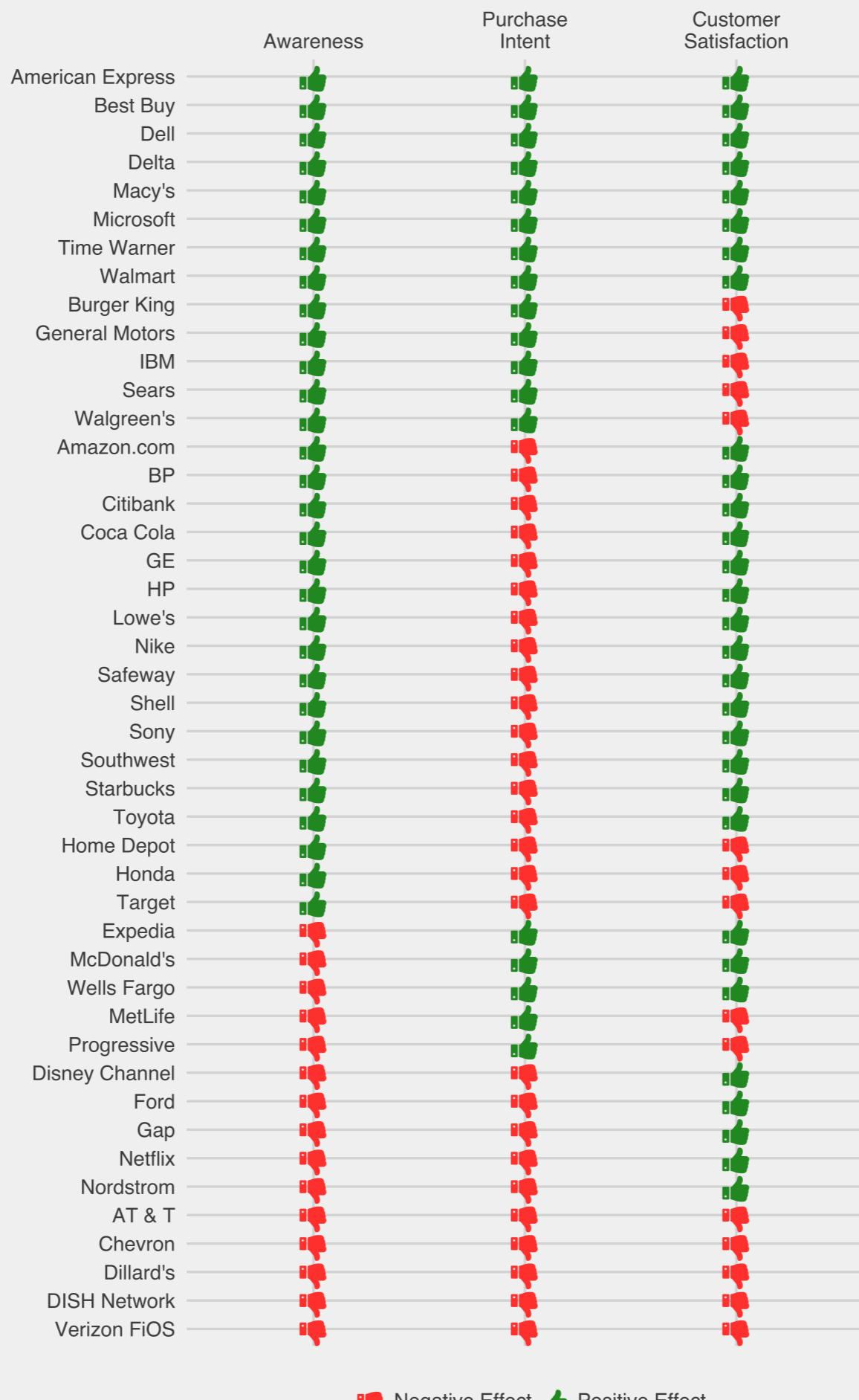


ggplot2

Prof. Ashwin Malshe

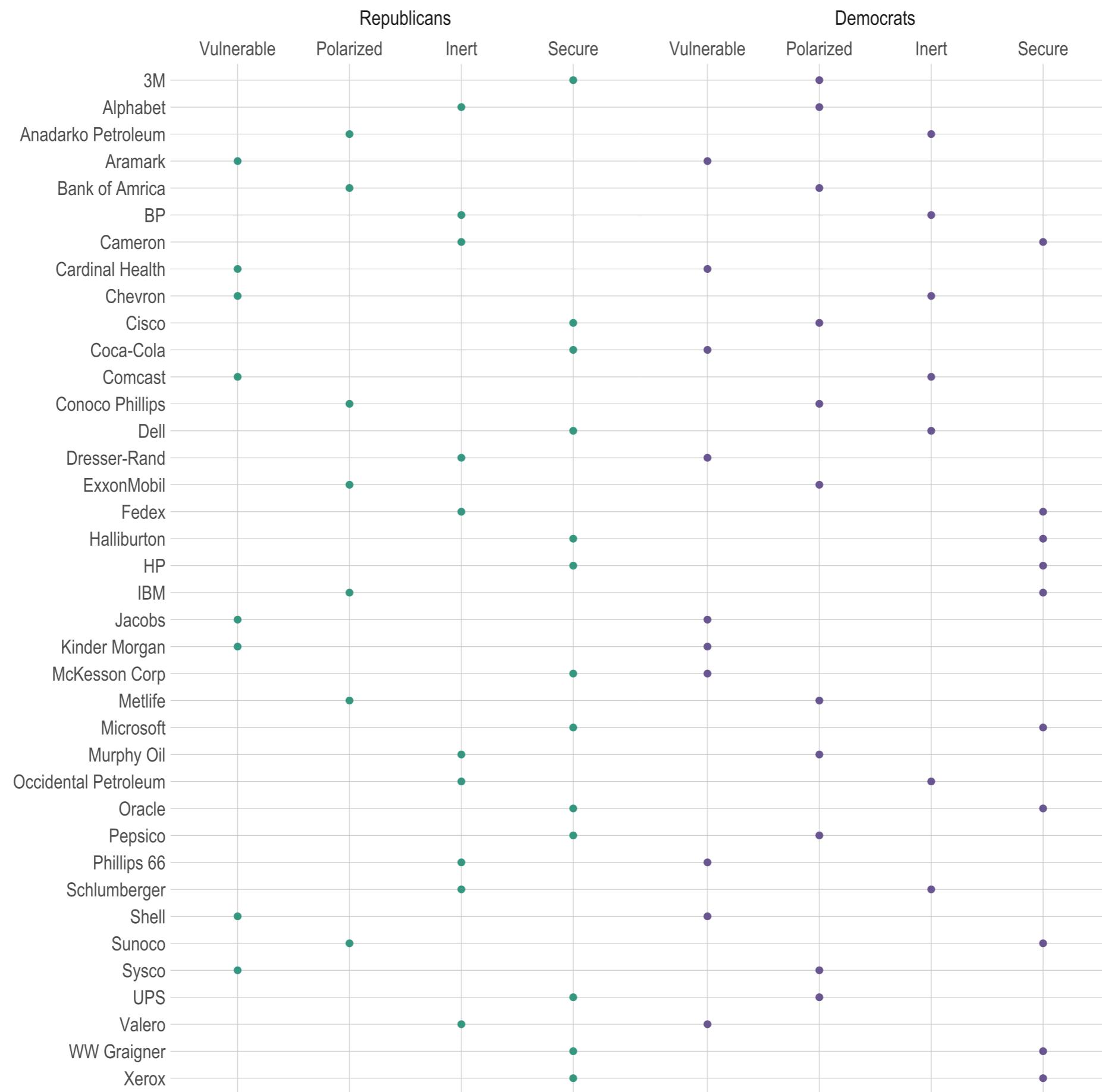
A Few Examples from My Work

Effectiveness of Owned Social Media

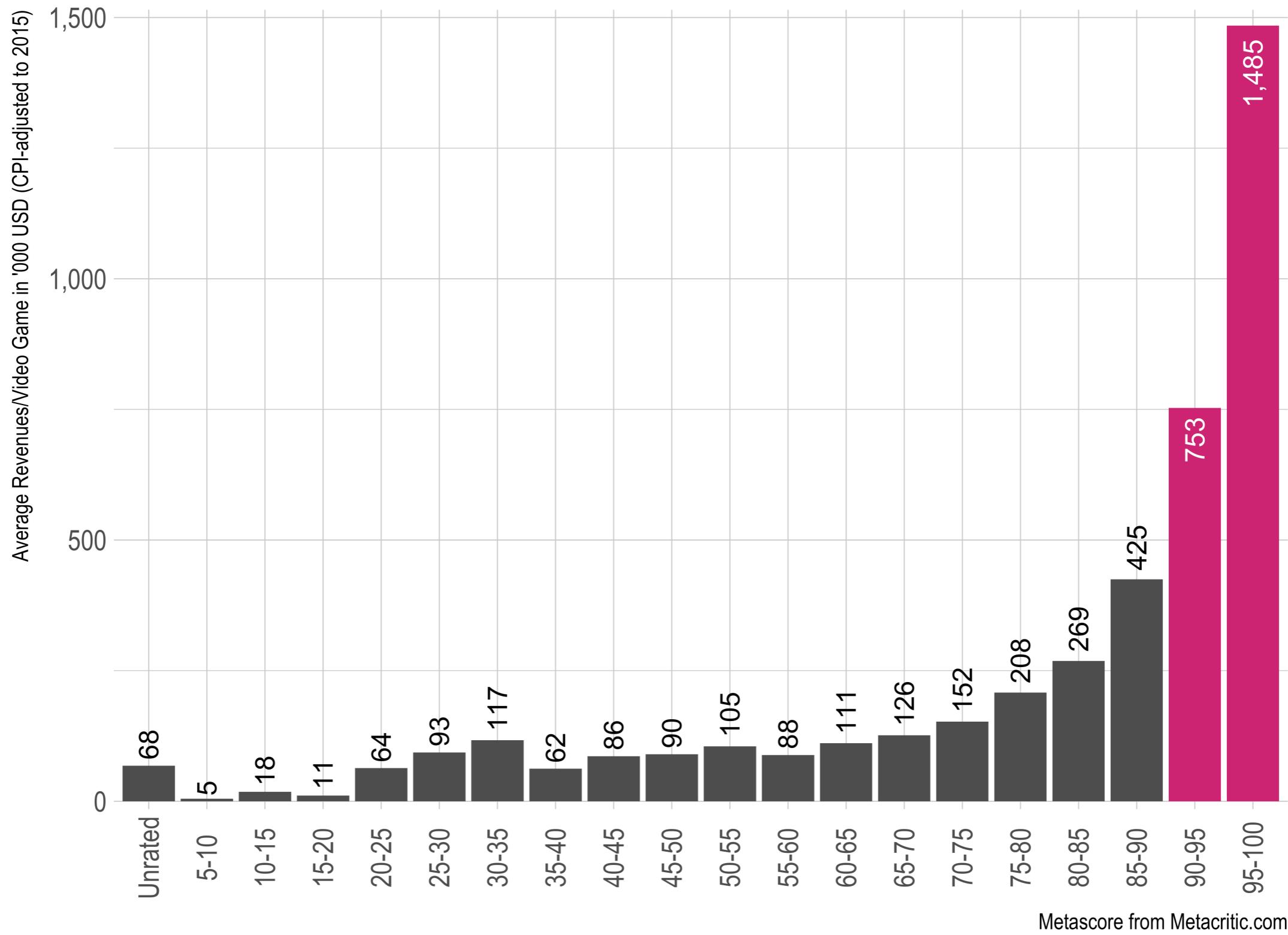


Malshe, Pauwels, Colicev 2019



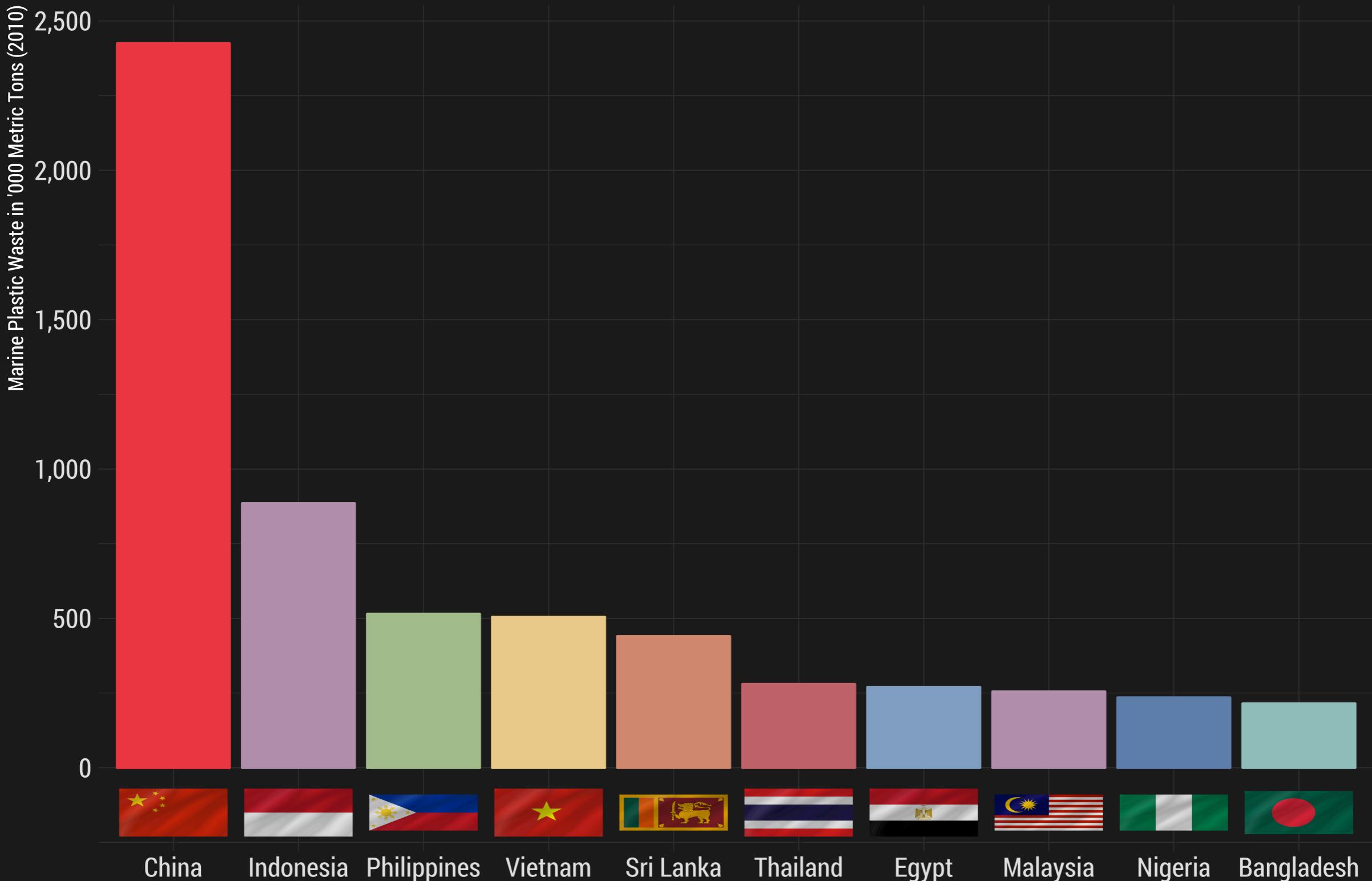


**Mittal, Malshe,
and Sridhar 2018**



Gretz, Malshe, Bauer, and Basuroy 2019

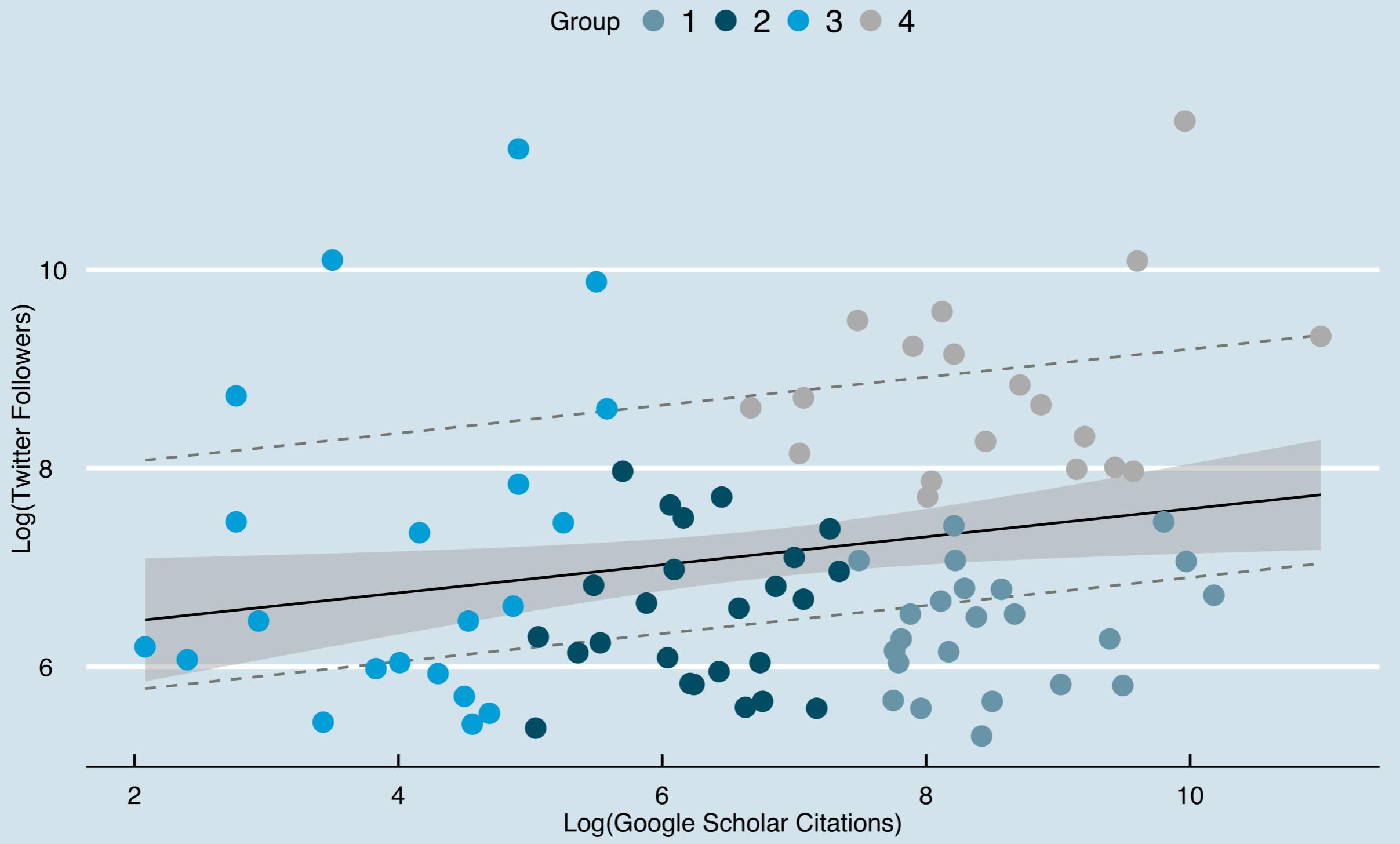
Top 10 Nations to Dump Plastic in the Oceans



© Ashwin Malshe 2019

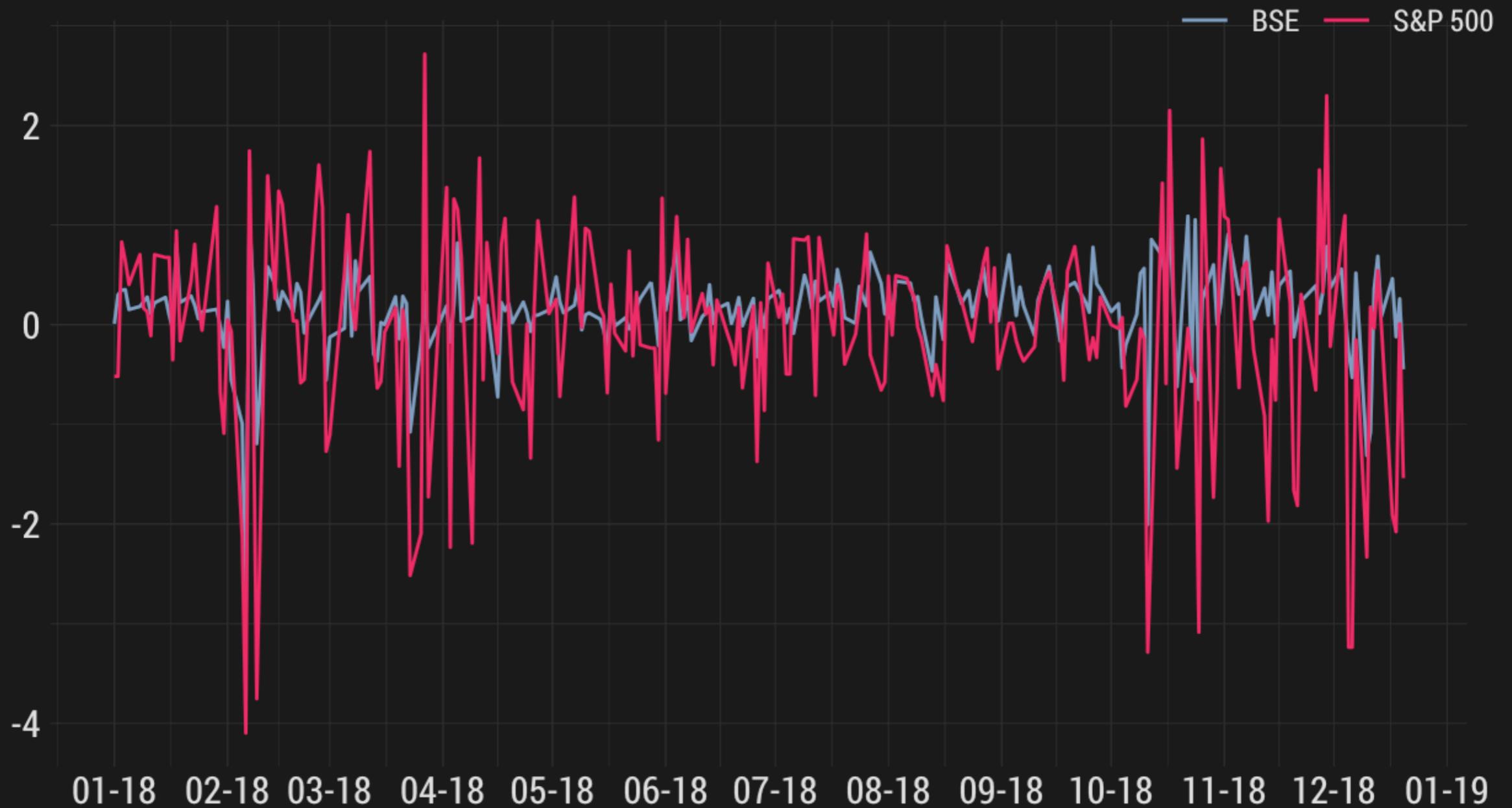
Source: Jambeck et al. (2015), 'Plastic waste inputs from land into the ocean,' Science

Google Scholar Citations and Twitter Followers



Daily Returns in 2018

BSE returns from close of day_t to open of day_{t+1} ($r_{xy} = 0.61$)



Grammar of Graphics

- The grammar of graphics is an answer to a question: what is a statistical graphic?
- The **layered** grammar of graphics focussing on the primacy of layers and adapting it for embedding within R
- The grammar tells us that a statistical graphic is a mapping from data to aesthetic attributes (color, shape, size) of geometric objects (points, lines, bars)
- The plot may also contain statistical transformations of the data and is drawn on a specific coordinate system.
- Faceting can be used to generate the same plot for different subsets of the dataset.
- It is the combination of these independent components that make up a graphic.

Components of Layered Grammar

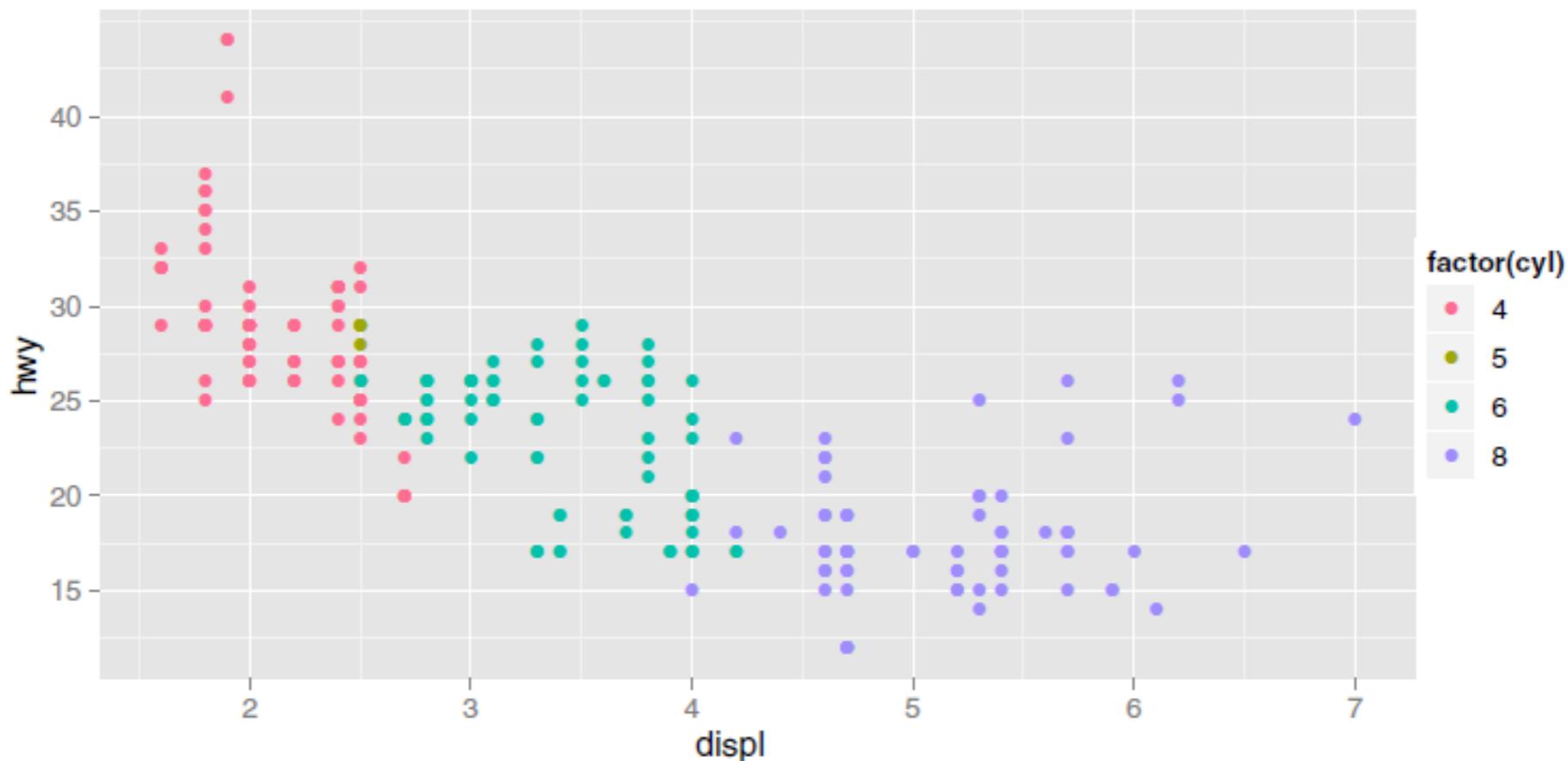
- Data
 - Layers
 - Scales
 - Coordinate system
 - Facets
 - Theme
- 
- We focus on these today**

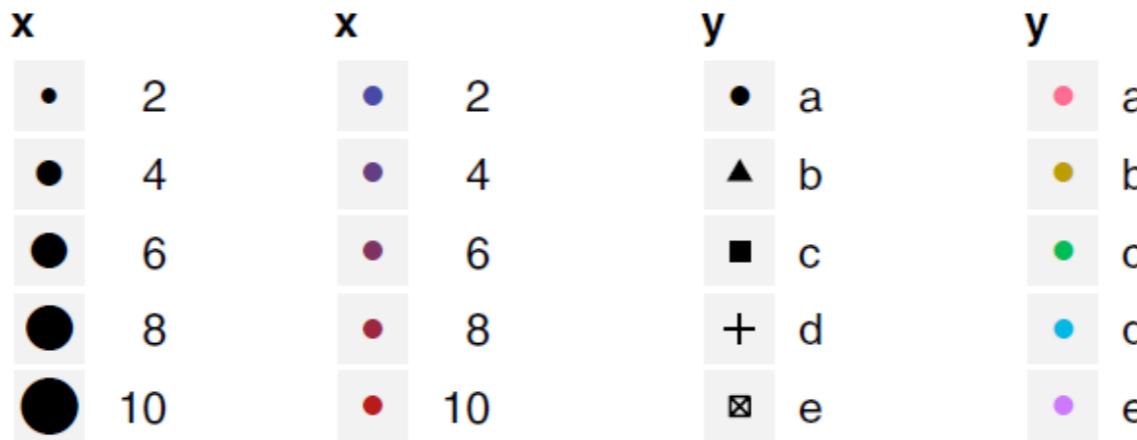
Layers

- Layers are responsible for creating the objects that we perceive on the plot
- A layer is composed of five parts:
 - Data
 - Aesthetic mapping
 - Statistical transformation (stat)
 - Geometric object (geom)
 - Position adjustment

A quick note on aesthetics mapping

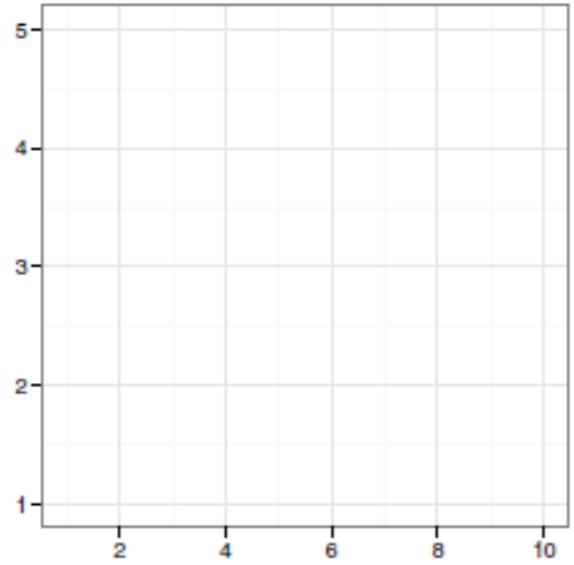
- Aesthetics are the properties that can be perceived on the graphic. Each aesthetic can be mapped to a variable, or set to a constant value.



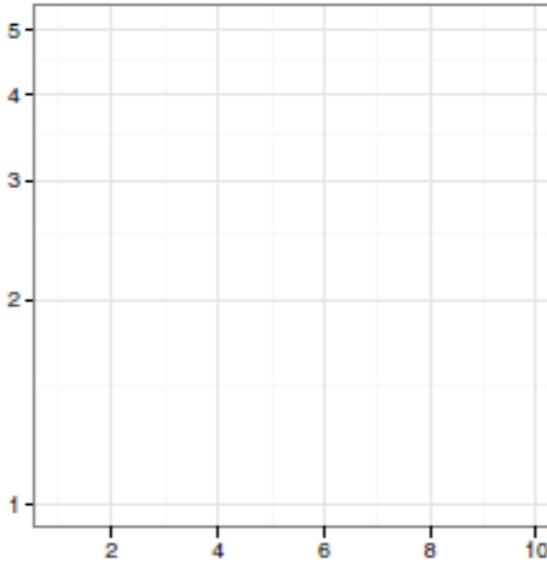


Scales

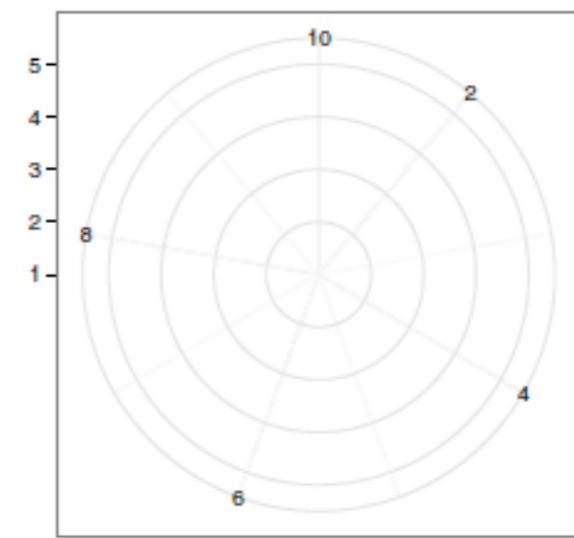
- A **scale** controls the mapping from data to aesthetic attributes, and we need a scale for every aesthetic used on a plot.
- Each scale operates across all the data in the plot, ensuring a consistent mapping from data to aesthetics.



Cartesian



Semi-log



Polar

Coordinate system

- A coordinate system, or **coord** for short, maps the position of objects onto the plane of the plot
- Position is often specified by two coordinates (x, y)
- Coordinate systems control how the axes and grid lines are drawn

Faceting

- This makes it easy to create small multiples each showing a different subset of the whole dataset
- This is a powerful tool when investigating whether patterns hold across all conditions
- The faceting specification describes which variables should be used to split up the data, and whether position scales should be free or constrained

Geoms

Geometric objects, or geoms for short, perform the actual rendering of the layer, controlling the type of plot that you create.

For example, using a point geom will create a scatterplot, while using a line geom will create a line plot.

- Graphical primitives:
 - `geom_blank()`: display nothing. Most useful for adjusting axes limits using data.
 - `geom_point()`: points.
 - `geom_path()`: paths.
 - `geom_ribbon()`: ribbons, a path with vertical thickness.
 - `geom_segment()`: a line segment, specified by start and end position.
 - `geom_rect()`: rectangles.
 - `geom_polygon()`: filled polygons.
 - `geom_text()`: text.

- One variable:
 - Discrete:
 - `geom_bar()`: display distribution of discrete variable.
 - Continuous
 - `geom_histogram()`: bin and count continuous variable, display with bars.
 - `geom_density()`: smoothed density estimate.
 - `geom_dotplot()`: stack individual points into a dot plot.
 - `geom_freqpoly()`: bin and count continuous variable, display with lines.

- Two variables:
 - Both continuous:
 - `geom_point()`: scatterplot.
 - `geom_quantile()`: smoothed quantile regression.
 - `geom_rug()`: marginal rug plots.
 - `geom_smooth()`: smoothed line of best fit.
 - `geom_text()`: text labels.
 - Show distribution:
 - `geom_bin2d()`: bin into rectangles and count.
 - `geom_density2d()`: smoothed 2d density estimate.
 - `geom_hex()`: bin into hexagons and count.

- At least one discrete:
 - `geom_count()`: count number of point at distinct locations
 - `geom_jitter()`: randomly jitter overlapping points.
- One continuous, one discrete:
 - `geom_bar(stat = "identity")`: a bar chart of precomputed summaries.
 - `geom_boxplot()`: boxplots.
 - `geom_violin()`: show density of values in each group.

- One time, one continuous
 - `geom_area()`: area plot.
 - `geom_line()`: line plot.
 - `geom_step()`: step plot.
 - Display uncertainty:
 - `geom_crossbar()`: vertical bar with center.
 - `geom_errorbar()`: error bars.
 - `geom_linerange()`: vertical line.
 - `geom_pointrange()`: vertical line with center.
 - Spatial
 - `geom_map()`: fast version of `geom_polygon()` for map data.
- Three variables:
 - `geom_contour()`: contours.
 - `geom_tile()`: tile the plane with rectangles.
 - `geom_raster()`: fast version of `geom_tile()` for equal sized tiles.

Geoms in ggplot2

| Name | Description |
|------------|-----------------------------------------------------------------------------|
| abline | Line, specified by slope and intercept |
| area | Area plots |
| bar | Bars, rectangles with bases on y-axis |
| blank | Blank, draws nothing |
| boxplot | Box-and-whisker plot |
| contour | Display contours of a 3d surface in 2d |
| crossbar | Hollow bar with middle indicated by horizontal line |
| density | Display a smooth density estimate |
| density_2d | Contours from a 2d density estimate |
| errorbar | Error bars |
| histogram | Histogram |
| hline | Line, horizontal |
| interval | Base for all interval (range) geoms |
| jitter | Points, jittered to reduce overplotting |
| line | Connect observations, in order of x value |
| linerange | An interval represented by a vertical line |
| path | Connect observations, in original order |
| point | Points, as for a scatterplot |
| pointrange | An interval represented by a vertical line, with a point in the middle |
| polygon | Polygon, a filled path |
| quantile | Add quantile lines from a quantile regression |
| ribbon | Ribbons, y range with continuous x values |
| rug | Marginal rug plots |
| segment | Single line segments |
| smooth | Add a smoothed condition mean |
| step | Connect observations by stairs |
| text | Textual annotations |
| tile | Tile plot as densely as possible, assuming that every tile is the same size |
| vline | Line, vertical |

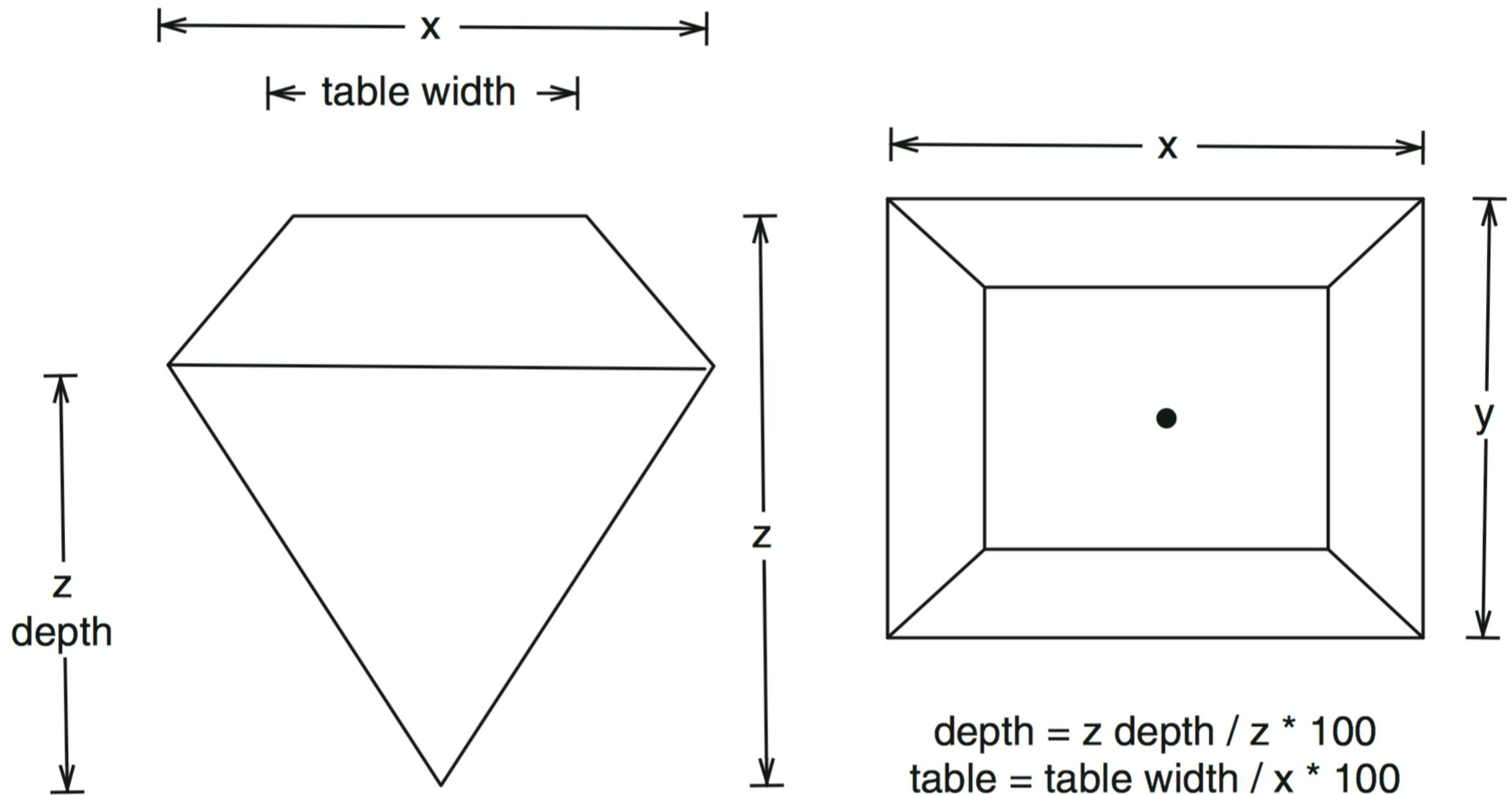


Fig. 3.1 How the variables x, y, z, table and depth are measured