ISA 401/501: Business Intelligence & Data Visualization

20: Charts for High Dimensional Data

Fadel M. Megahed, PhD

Professor of Information Systems and Business Analytics
Farmer School of Business
Miami University

@FadelMegahed
fmegahed
fmegahed@miamioh.edu
Automated Scheduler for Office Hours

Fall 2024

Learning Objectives for Today's Class

- Describe what is high dimensional data.
- Provide some examples for graphs used for high dimensional datasets.
- Construct these graphs using software

High Dimensional Data

What Do we Mean by High Dimensional Data?

03:00

Activity

> In 3 minutes, define the terms in next tab in the context of this table.

Order ID	Order Date	Order Priority	Product Container	Product Cost	Ship Date
1	1/1/2022	5 - low	Large box	25	1/5/2022
2	1/4/2022	4 - not specified	Small Box	36	1/7/2022
3	1/15/2022	2- high	Small Box	38	1/17/2022
3	1/15/2022	2- high	Small Box	41	1/17/2022
3	1/15/2022	2- high	Jumbo Box	44	1/17/2022
3	1/15/2022	2- high	Wrap Bag	33	1/17/2022
4	1/18/2022	1- urgent	Small Box	33	1/19/2022

Showing 1 to 7 of 11 entries

Previous 1 2 Next

Taxonomy

Based on the number of attributes:

- 1: Univariate
- 2: Bivariate
- 3: Trivariate
- 4+: Multivariate

Things to Think about:

- What is the problem with visualizing multivariate (especially when $p > 6 - 7$ dimensions) data? __
- Any ideas about what to do? __

Examples of High Dimensional Charts

Hans Rosling: The Best Stats You Have Seen

Activity | Your Solution

> While watching this video, please answer the questions in the next tab!!

Hans Rosling: The Best Stats You Have Seen

Activity | Your Solution

- What data is represented in this visualization? Be specific. __
- How is each data type visually encoded? _
- Do you think the encodings are appropriate? _

So What is the Motion Bubble Chart?

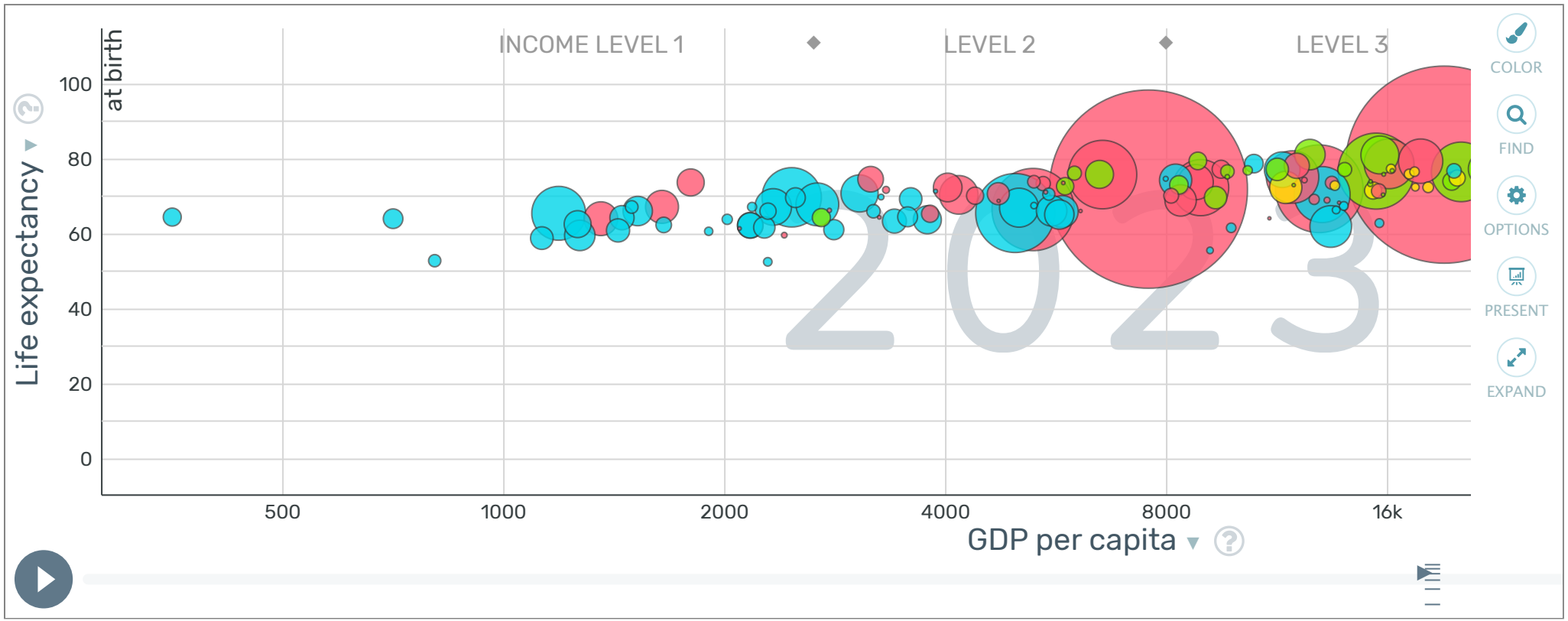
> Motion charts are essentially **animated bubble charts**. A bubble chart shows data using the **x-axis**, **y-axis**, and the **size** and **color** of the bubble. A motion chart displays **changes over time by showing movement within the two-dimensional space and changes in the size and color of the bubbles**. -- Juice Analytics

Encoding mechanisms:

- **x-axis** is typically used to encode a **numeric variable**
- **y-axis** is also used to encode a **numeric variable**
- **area** is used to encode a **numeric/ordinal variable**
- **color** is typically used to encode a **nominal variable**
- **motion/animation** is typically used to encode **time**

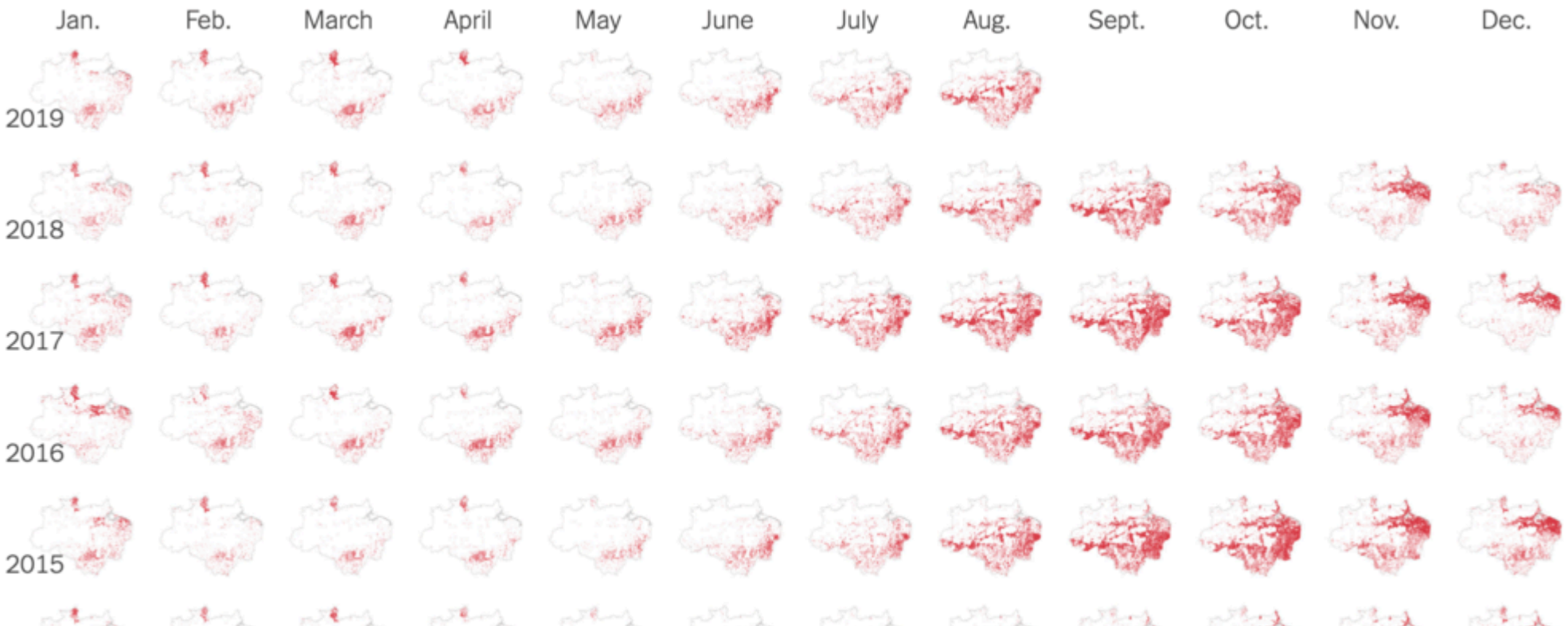
Live Demo: Creating Bubble Charts in Power BI

Let us use Power BI to create a similar chart to the one created by Hans Rosling. We will be using the gapminder.csv.



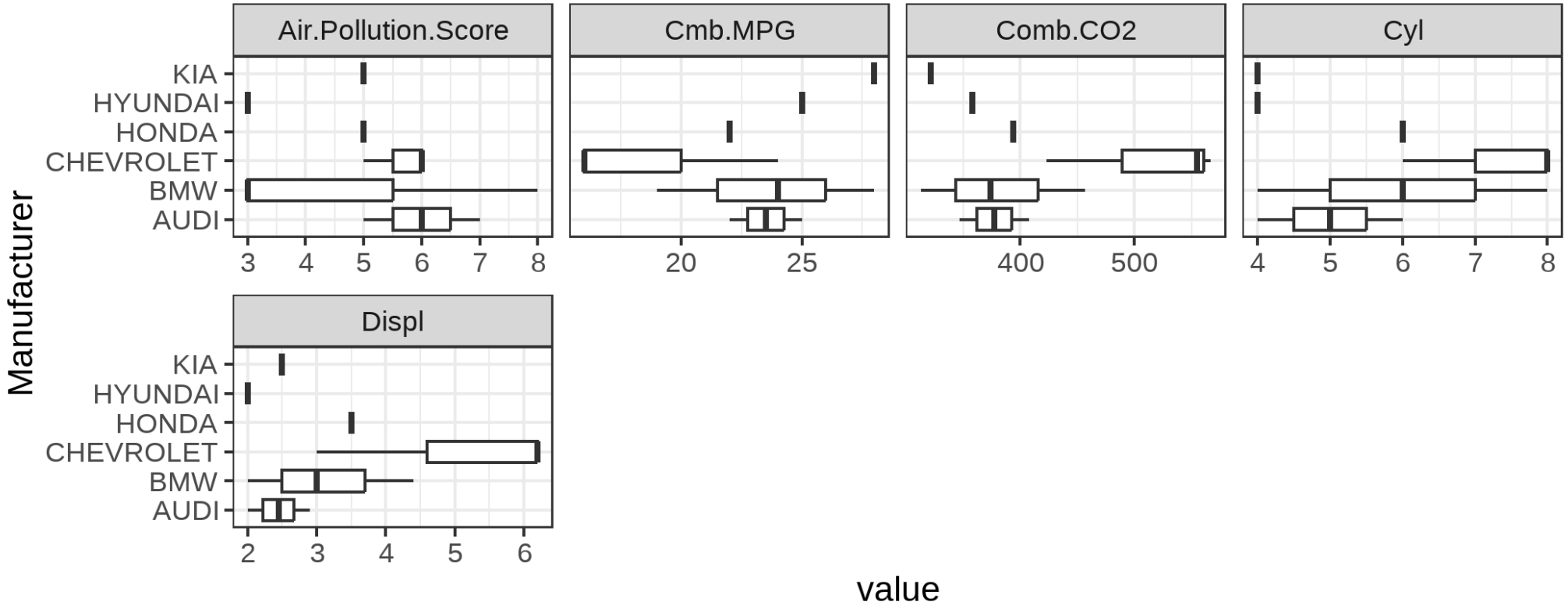
Small Multiples

> Illustrations of postage-stamp size are indexed by category or a label, sequenced over time like the frames of a movie, or ordered by a quantitative variable not used in the single image itself -- Tufte, E.R.: Envisioning Information, Graphics Press, 1990



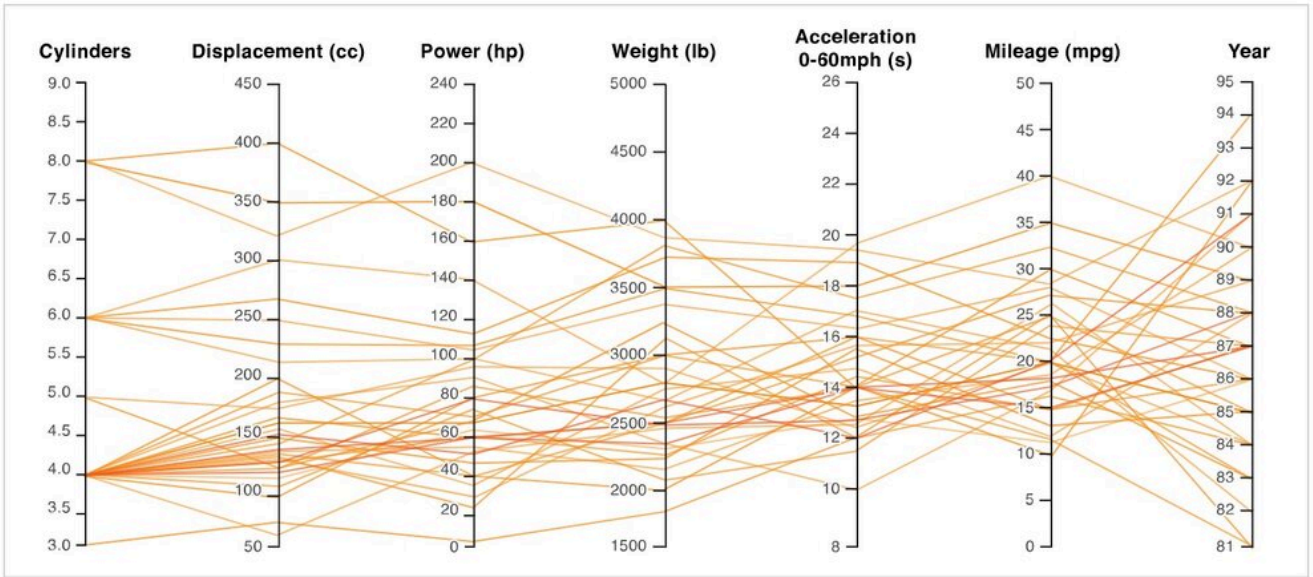
Small Multiples in Power BI

Let us use Power BI to create a similar chart to the one below. We will be using the mpg_2023_large.csv.



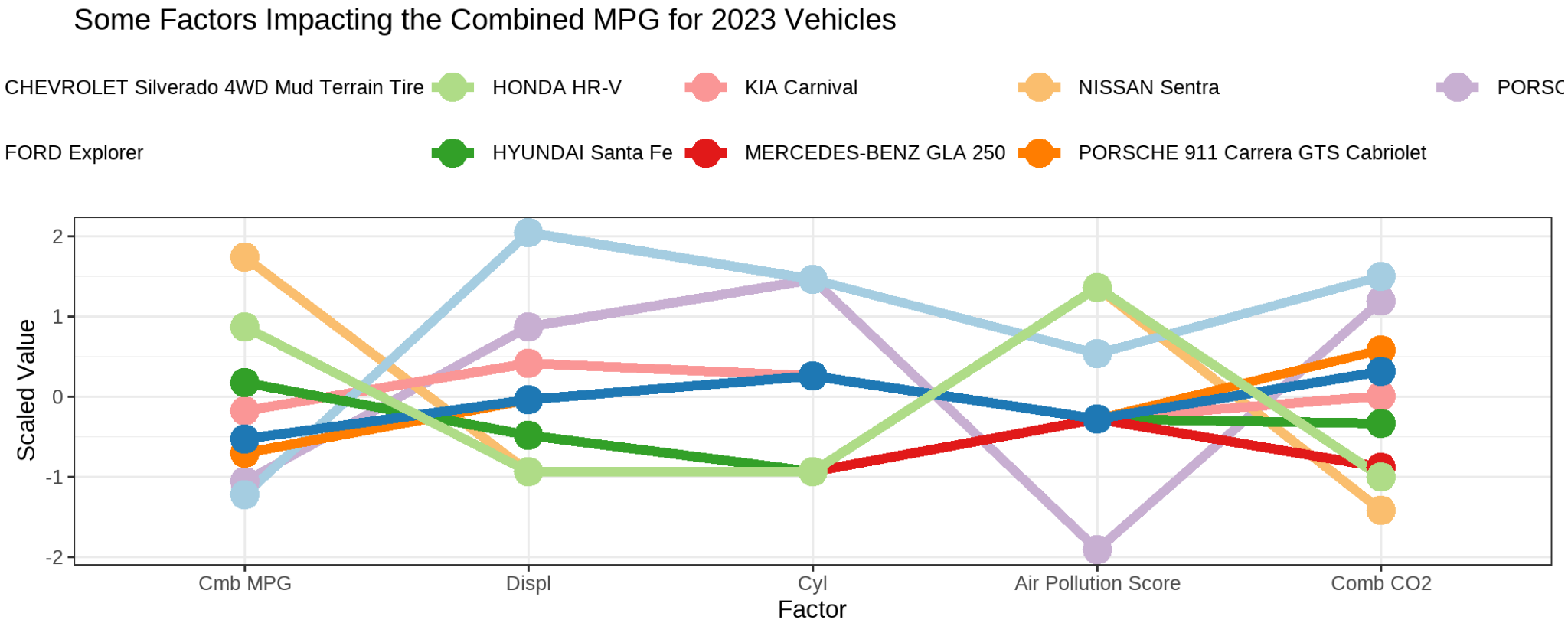
Parallel Coordinates

> Parallel coordinates is a visualization technique used to plot individual data elements across many performance measures. Each of the measures corresponds to a vertical axis and each data element is displayed as a series of connected points along the measure/axes -- Juice Analytics' Defintion



Parallel Coordinates in Power BI

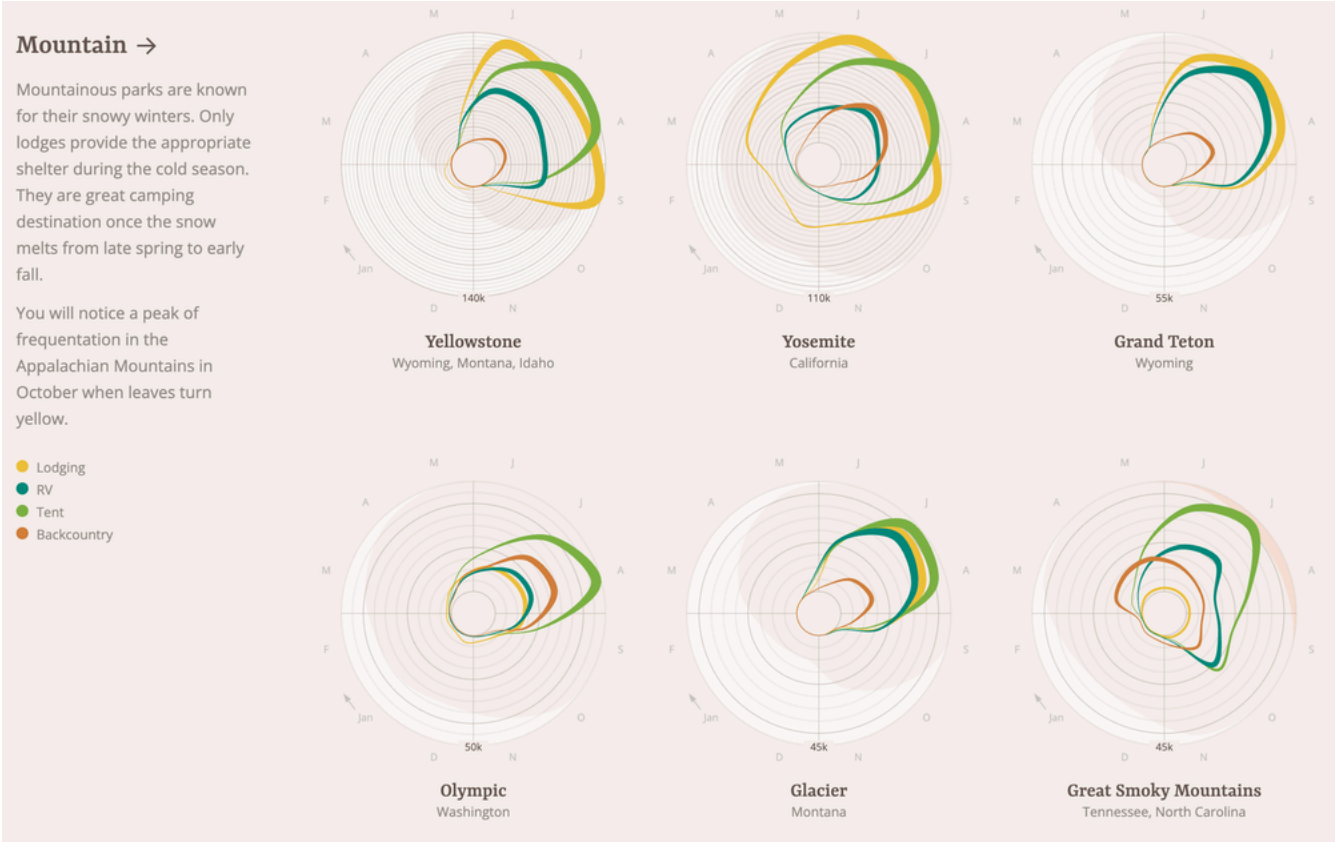
Let us visualize the mpg_2023_sample.csv using a parallel coordinates plot in Power BI.



Data source: https://www.fueleconomy.gov/feg/download.shtml

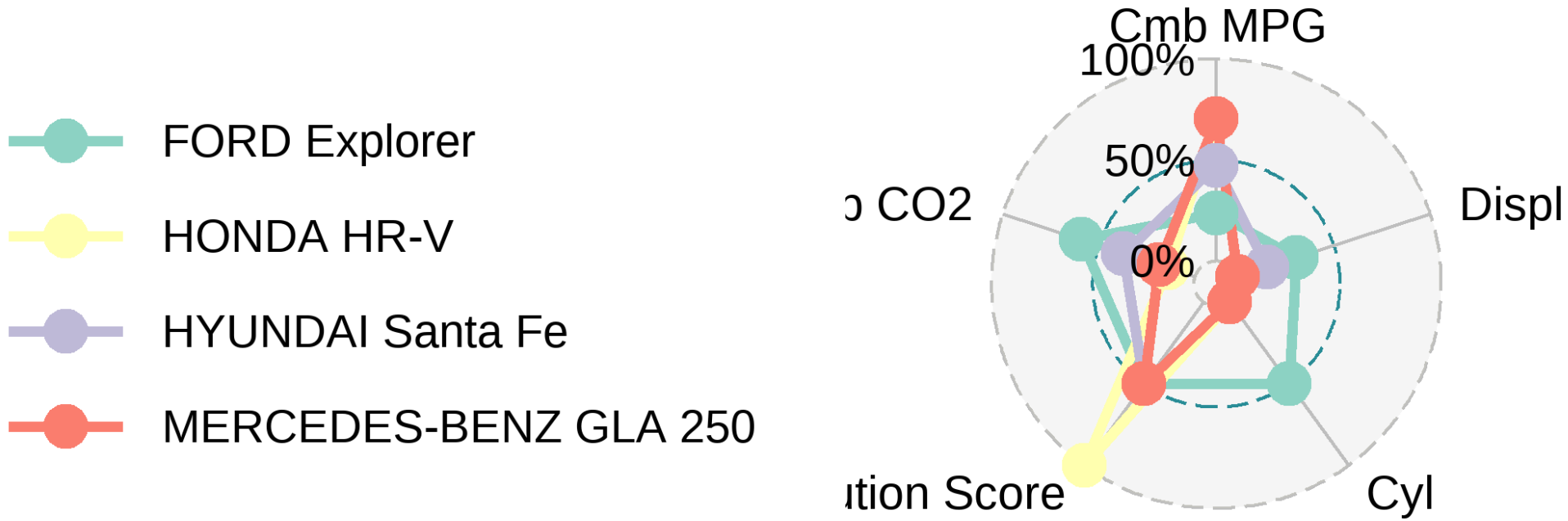
Radar Charts

> Charts show how individual things perform across multiple measures



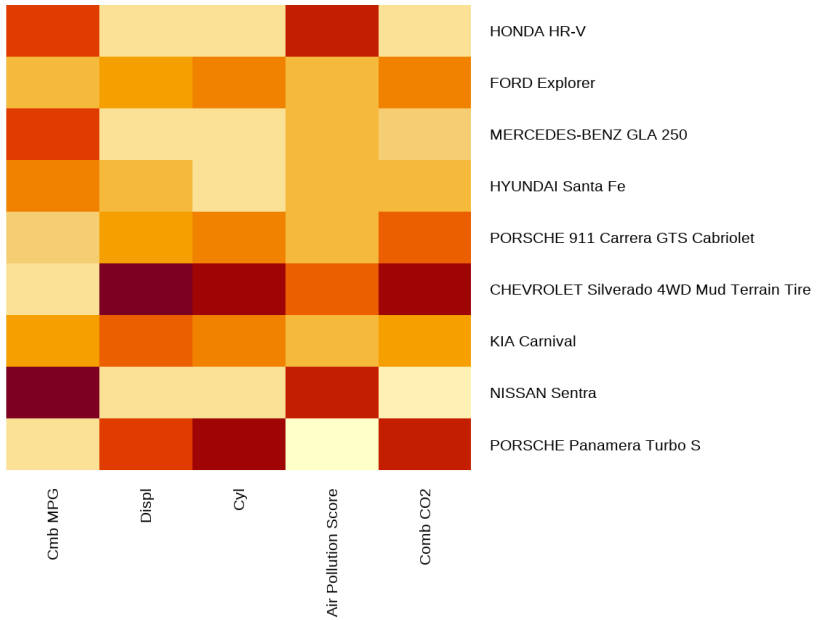
Radar Charts in Power BI

Let us add the Radar Chart App to our Power BI and use it to visualize the mpg_2023_sample.csv.



Other Charts: HeatMap

- each column is a variable
- each obs is a row
- each square is a value; closer to yellow the higher



Other Charts: TreeMaps

> Treemaps simultaneously show the big picture, comparisons of related items, and allow easy navigation to the details. Juice Analytics

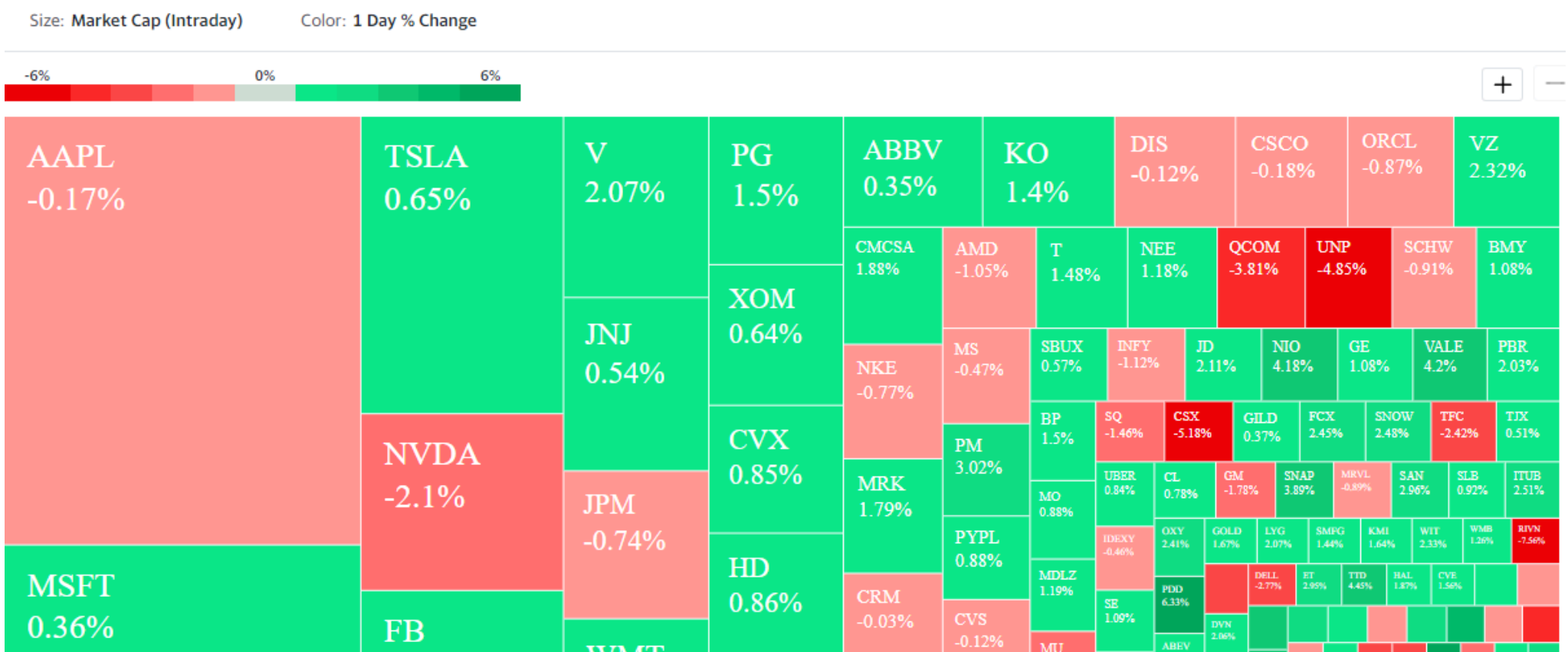
Encoding mechanisms:

Each box in a treemap can show two measures:

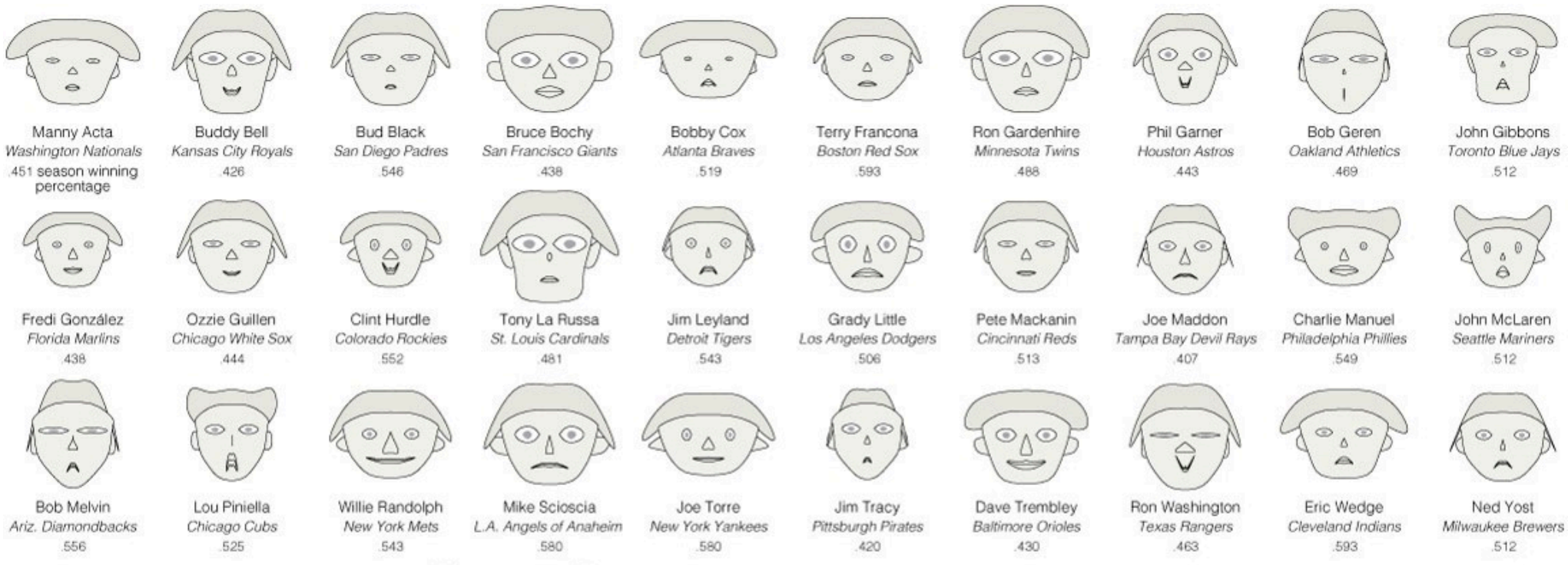
- **area of the boxes** should be a **quantity measure**. The measures should sum up along the hierarchical structure of the data. The sum of all the elements in one branch need to sum to the value of the branch as a whole.
- **Color of the boxes** is best suited to a **measure of performance or change** such as growth over time, average conversion rate, or customer satisfaction.

Other Charts: TreeMaps

> Treemaps simultaneously show the big picture, comparisons of related items, and allow easy navigation to the details. Juice Analytics

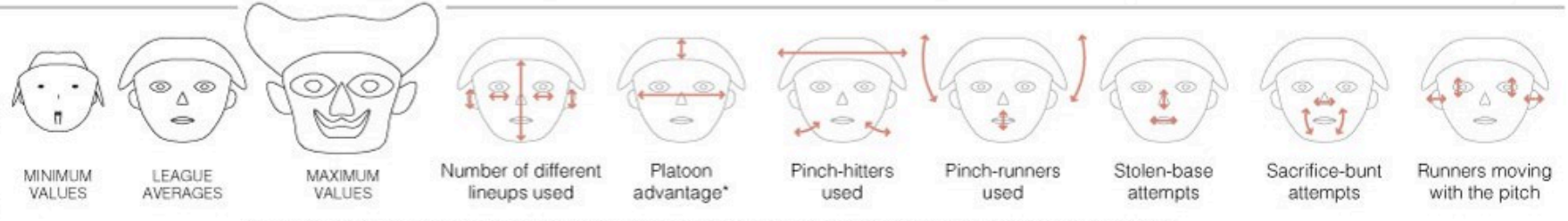


Other Charts: Chernoff Faces



SMILE IF YOU BUNT

Steve C. Wang, an associate professor of statistics at Swarthmore College, charted baseball managers from the 2007 season as Chernoff faces, a method of using the heights, widths and angles of facial features to represent different sets of numbers.



*Percentage of players who had the advantage of batting against an opposite-handed pitcher at the start of the game.
Note: Because different rules cause National League managers to use more pinch-hitters, for example, each manager's rates are compared with his league's average.

JONATHAN CORUM/ THE NEW YORK TIMES

Recap

Summary of Main Points

- Describe what is high dimensional data.
- Provide some examples for graphs used for high dimensional datasets.
- Construct these graphs using software

Non-graded Activity: Kahoot

Let us go to Kahoot and compete for a $10 Starbucks gift card. To evaluate your understanding of the material, please answer the questions correctly and as quickly as possible to get the most points.