

ISA 401: Business Intelligence & Data Visualization

19: Charts for High Dimensional Data

Fadel M. Megahed, PhD

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

Twitter: [FadelMegahed](#)

GitHub: [fmegahed](#)

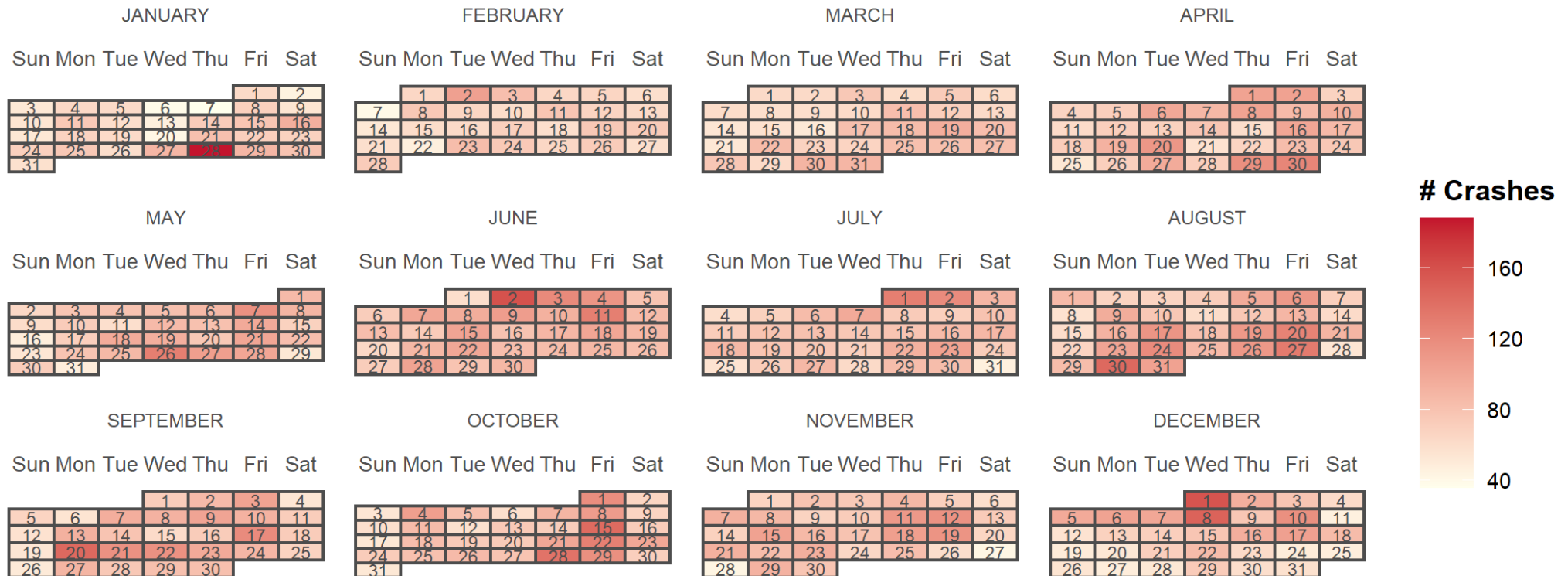
Email: fmegahed@miamioh.edu

Office Hours: [Automated Scheduler for Virtual Office Hours](#)

Spring 2022

Warm-up/Non-graded Activity: A Calendar Plot

Let us use  inside Power BI to create this based on [cincy_daily_crashes.csv](https://data.cincy.gov/ds734/cincy-daily-crashes.csv).



Created by: Fadel Megahed | Data source: City of Cincy Open Data Portal (rvmt-pkmq)

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

03:00

What Do we Mean by High Dimensional Data.

Activity

Your Solution

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

What Do we Mean by High Dimensional Data.

Activity

Your Solution

Data Types: (Edit below)

- Multivariate data: __
- Big Data: __
- Tall Data: __
- Wide Data: __
- High Dimensional Data: __

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?

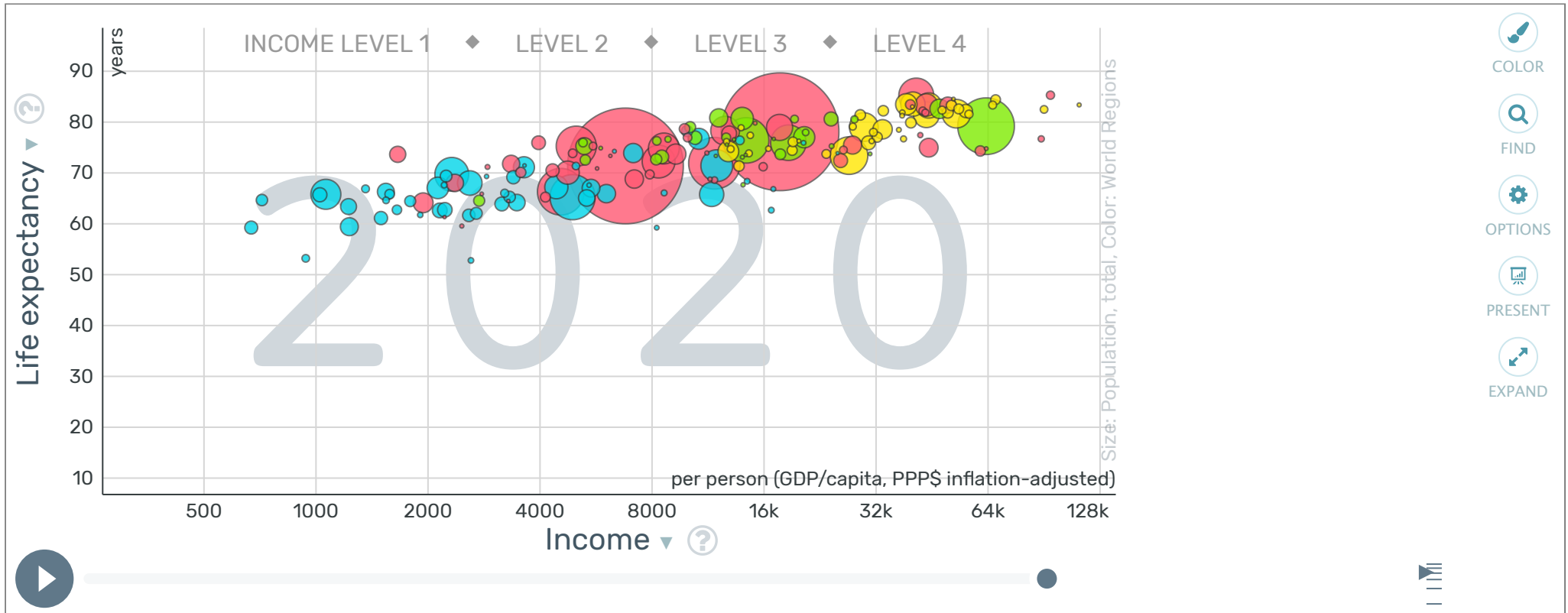
Motion charts are essentially **animated bubble charts**. A bubble chart shows data using the **x-axis**, **y-axis**, and the **size** and **color** of the bubble. A motion chart displays **changes over time by showing movement within the two-dimensional space and changes in the size and color of the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

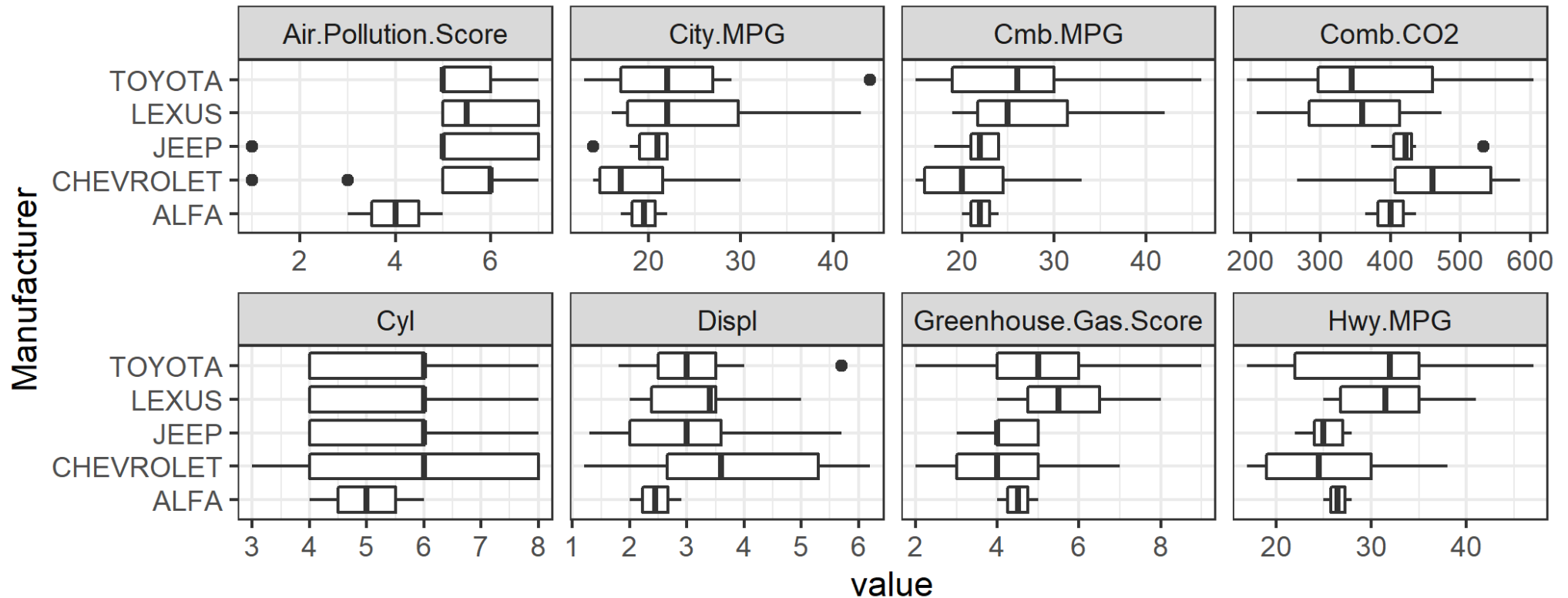


Small Multiples

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](#)

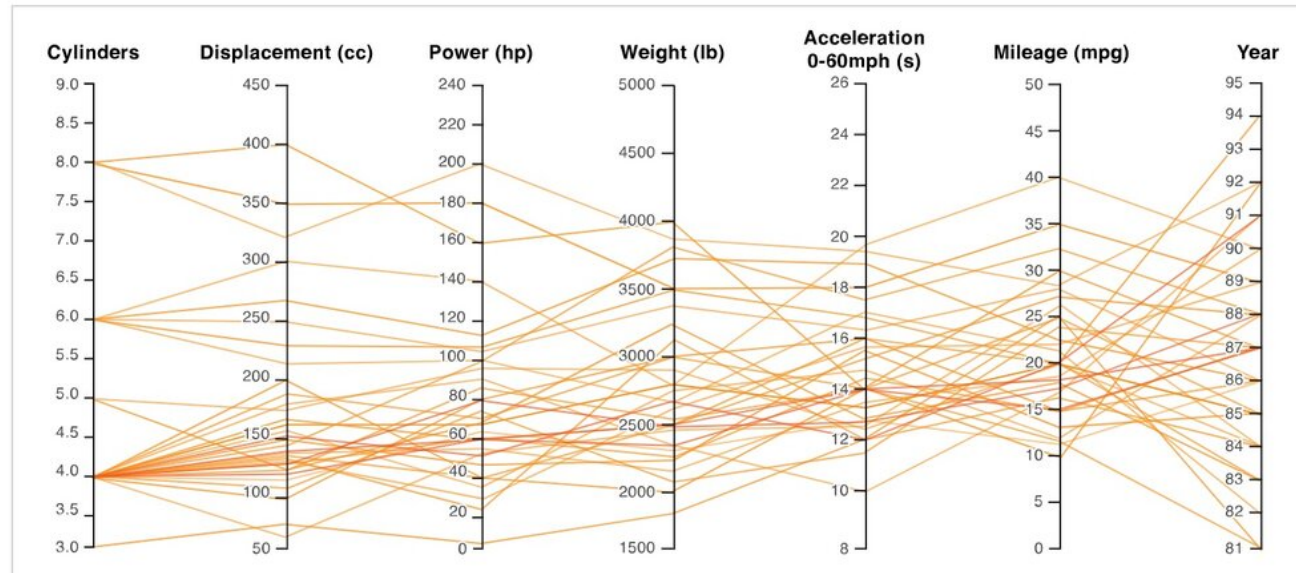
Small Multiples in Power BI

Let us use Power BI to create a similar chart to the one below. We will be using the [mpg_2022_large_sample.csv](#).



Parallel Coordinates

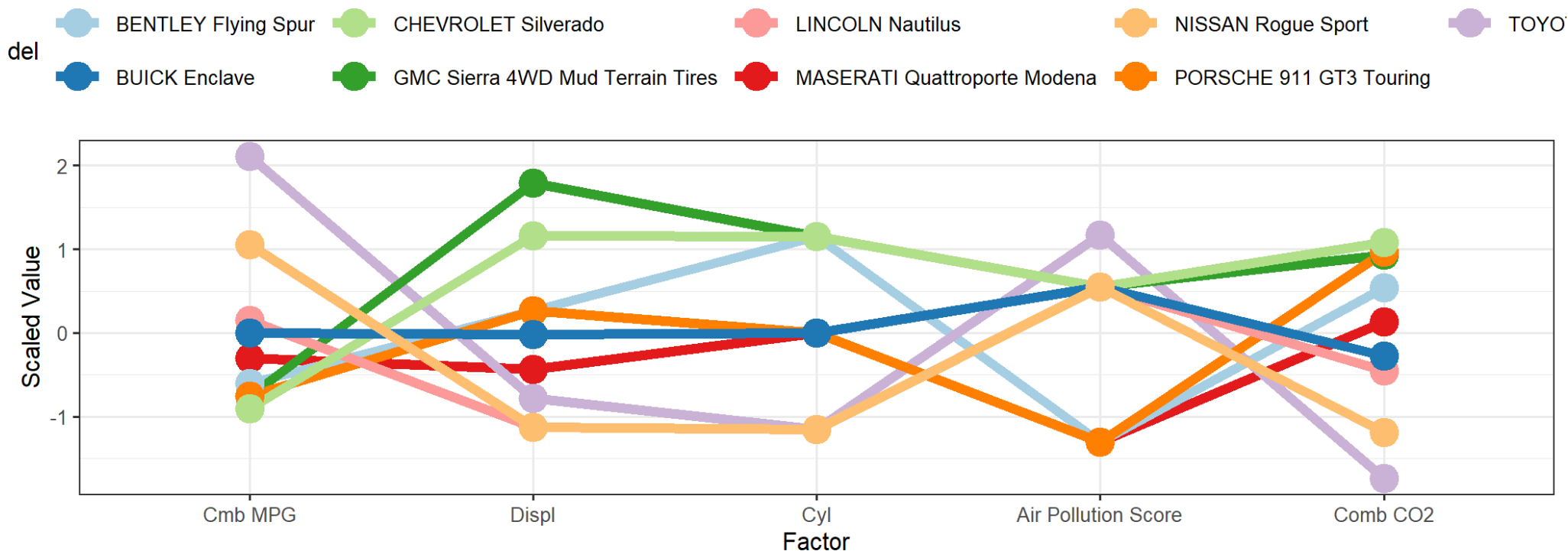
Parallel coordinates is a visualization technique used to plot individual data elements across many performance measures. Each of the measures corresponds to a vertical axis and each data element is displayed as a series of connected points along the measure/axes -- [Juice Analytics' Definition](#)



Parallel Coordinates in Power BI

Let us add the [Custom Parallel Coordinates Charting App](#) to our Power BI and use it to visualize the [mpg_2022_sample.csv](#).

Some Factors Impacting the Combined MPG for 2022 Vehicles



Radar Charts

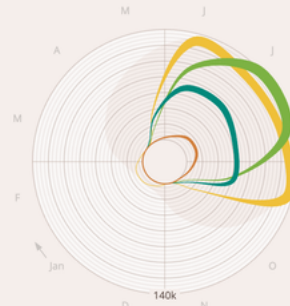
Charts show how individual things perform across multiple measures

Mountain →

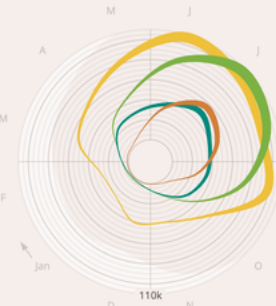
Mountainous parks are known for their snowy winters. Only lodges provide the appropriate shelter during the cold season. They are great camping destination once the snow melts from late spring to early fall.

You will notice a peak of frequentation in the Appalachian Mountains in October when leaves turn yellow.

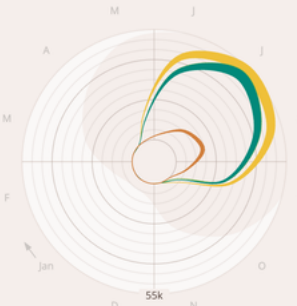
- Lodging
- RV
- Tent
- Backcountry



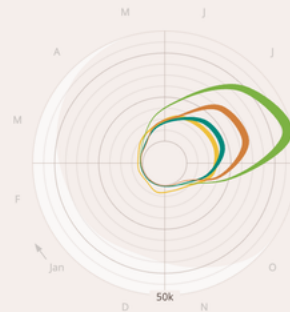
Yellowstone
Wyoming, Montana, Idaho



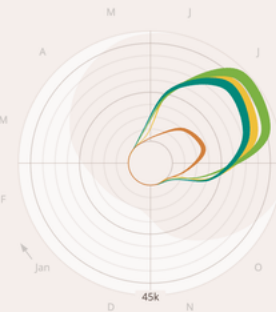
Yosemite
California



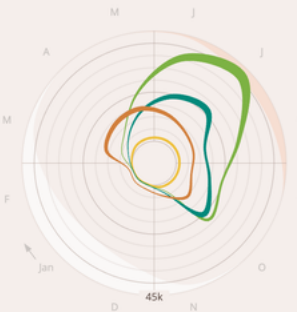
Grand Teton
Wyoming



Olympic
Washington



Glacier
Montana

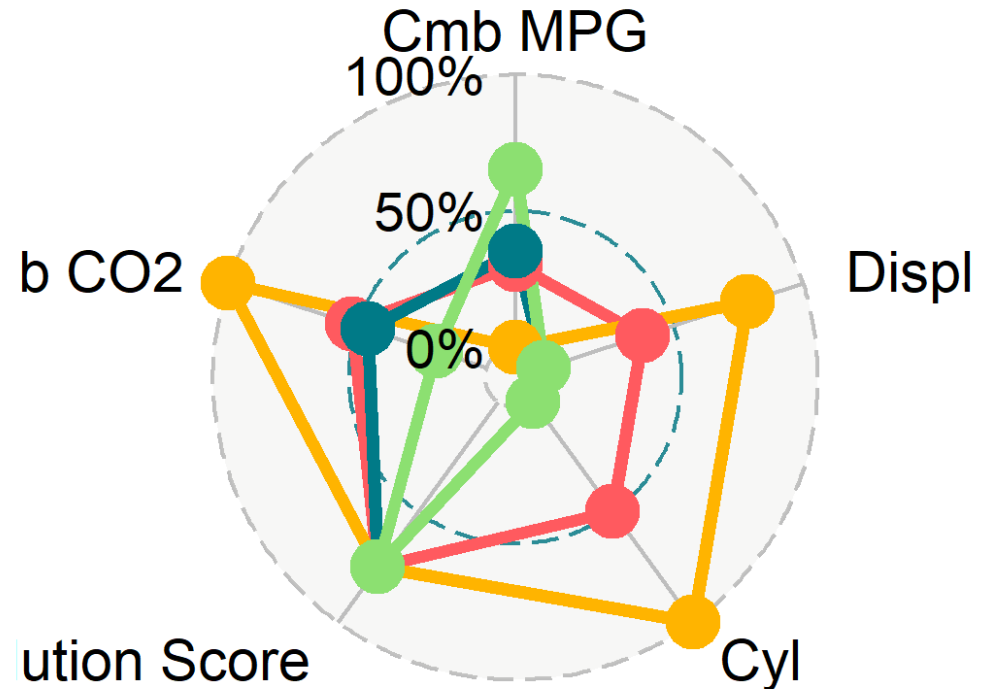


Great Smoky Mountains
Tennessee, North Carolina

Radar Charts in Power BI

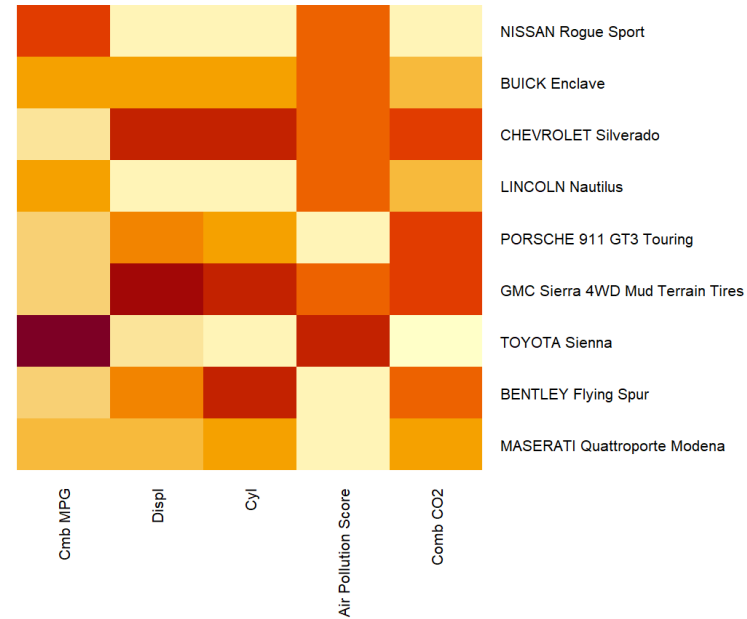
Let us add the [Radar Chart App](#) to our Power BI and use it to visualize the [mpg_2022_sample.csv](#).

- BUICK Enclave
- CHEVROLET Silverado
- LINCOLN Nautilus
- NISSAN Rogue Sport



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](#)

Other Charts: Chernoff Faces

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