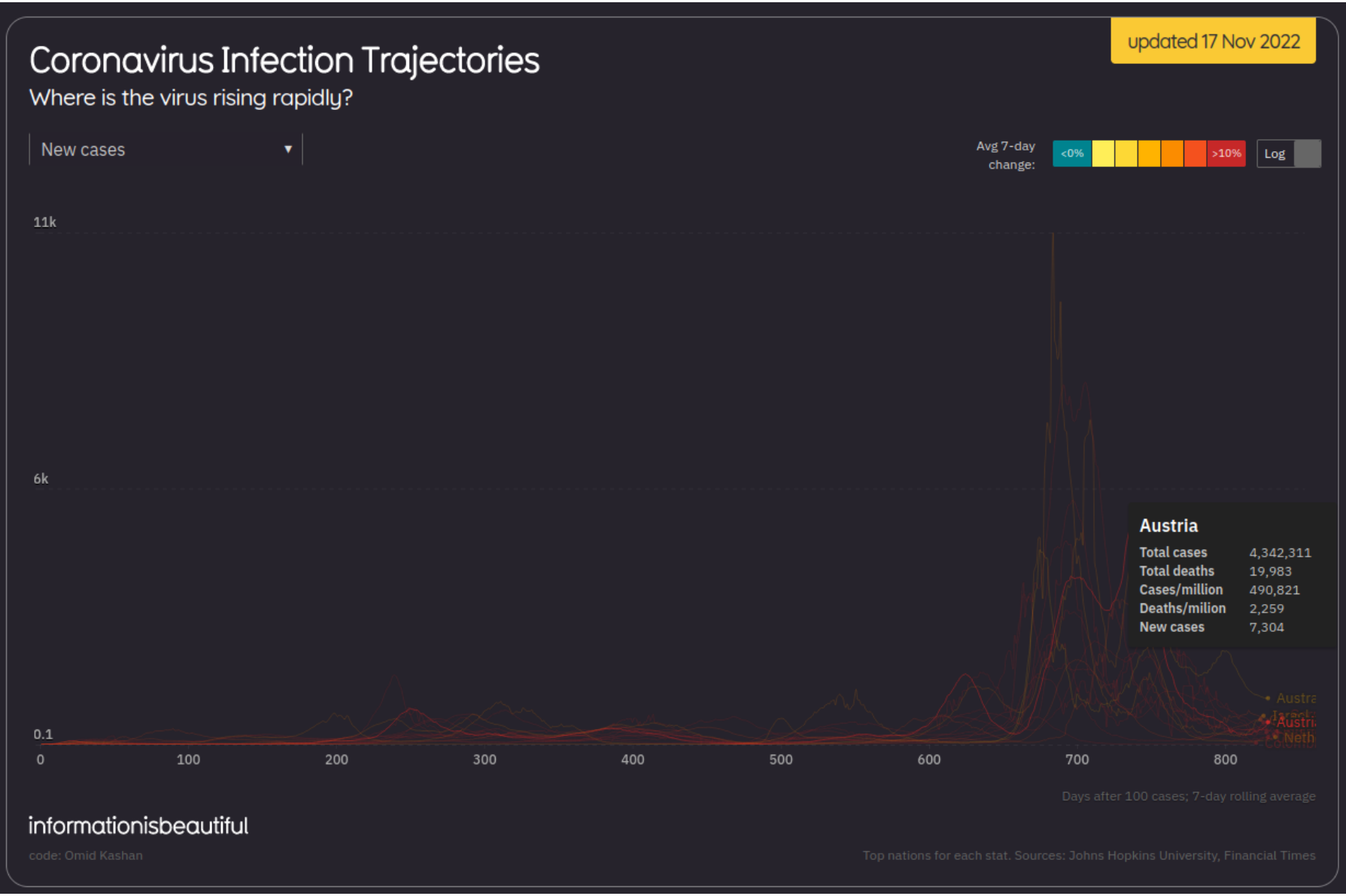
Interactive Data Exploration
with Plotscaper

Adam Bartonicek
23-10-2022

Interactive data visualizations are more and more popular...

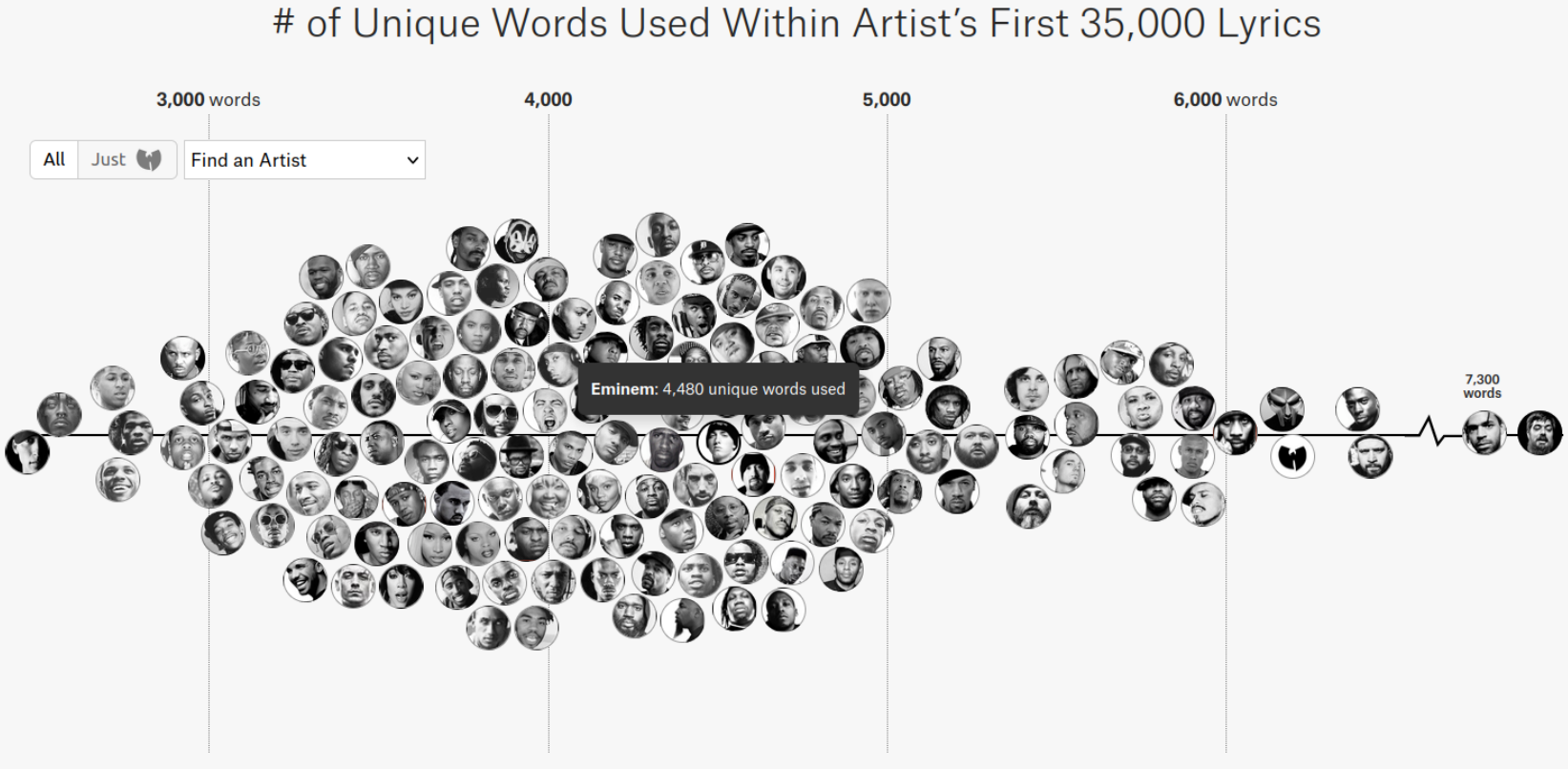

Coronavirus Infection Trajectories
Where is the virus rising rapidly?
updated 17 Nov 2022
New cases
Avg 7-day change:
<0%
>10%
Log
11k
6k
0.1
0
100
200
300
400
500
600
700
800
Austria
Total cases 4,342,311
Total deaths 19,983
Cases/million 490,821
Deaths/milion 2,259
New cases 7,304
Austra
Austri
Neth
Days after 100 cases; 7-day rolling average
informationisbeautiful
code: Omid Kashan
Top nations for each stat. Sources: Johns Hopkins University, Financial Times

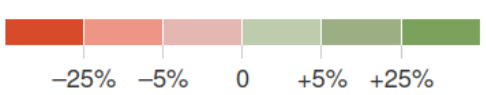
ThePudding

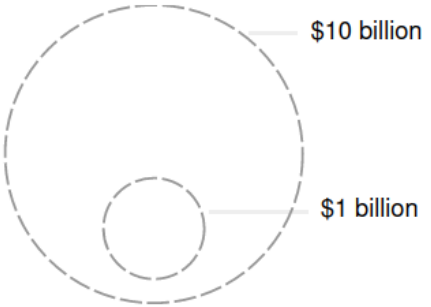
The Largest Vocabulary In Hip Hop

Rappers, ranked by the number of unique words used in their lyrics

By Matt Daniels





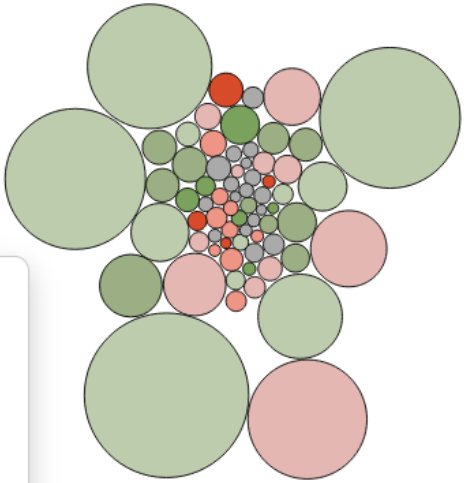


*The visusalisation works best in modern browsers.

By Harkanwal Singh
15 May, 2014 02:00 PM Quick Read

Social Security and Welfare

Education



Impairment of Debt Relating to Child Support
Amount: $323.0 million
% change: 9.12%
Department: Inland Revenue

Uncomfortable personal question

Do you use interactive visualizations to explore your data?

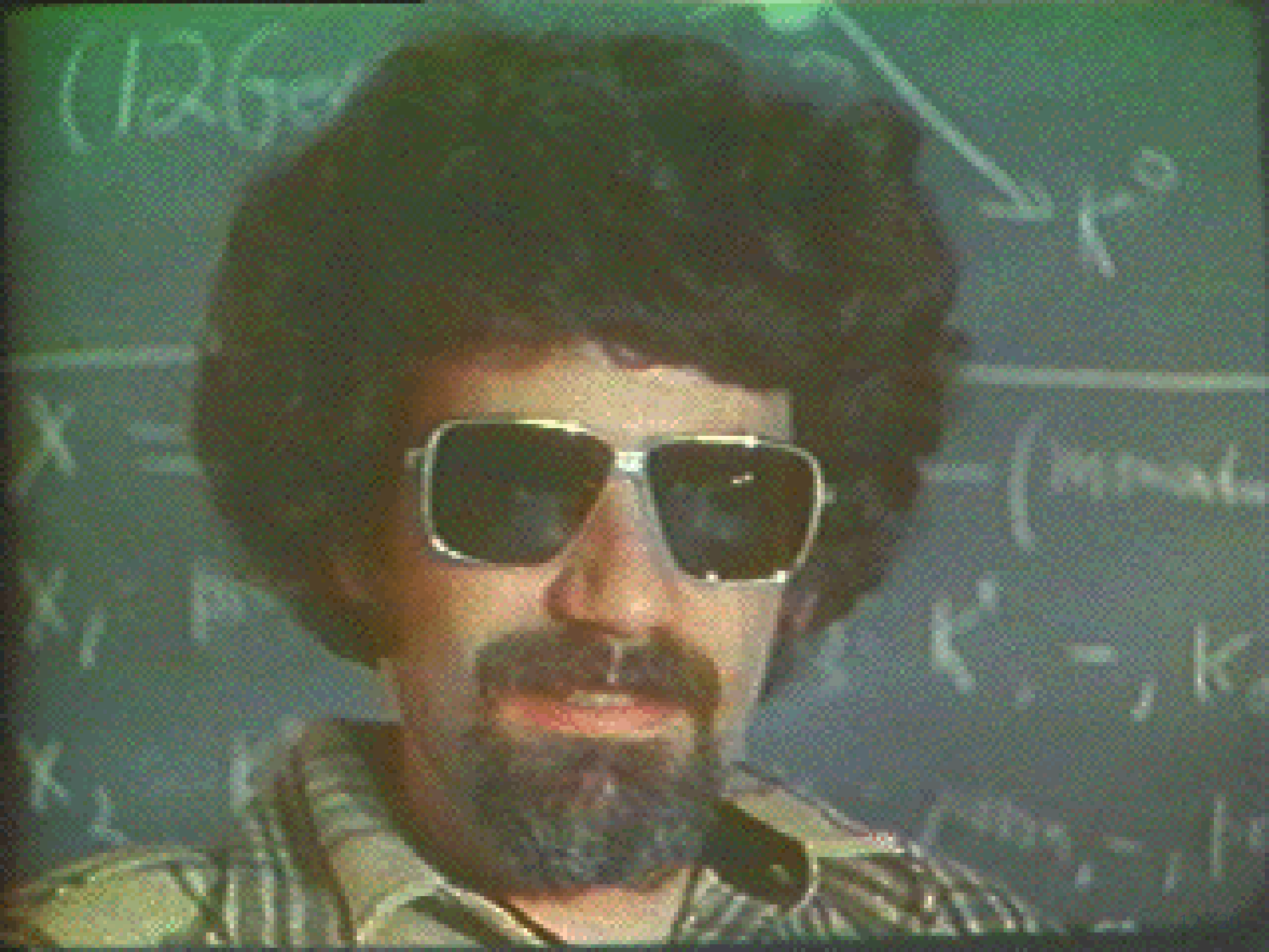
> "Rapid advances in the power, simplicity, and familiarity of visualization combined with an increasing awareness of the potential of visual communication have pushed the field to the cusp of mainstream breakthrough in society.
>
> [...] Still, very few data science tools and workflows actually employ interactive visualization as anything more than a mere communication tool used to disseminate results at the end of an investigation to stakeholders and the general public."
>
> (Batch & Elmquist, 2017)

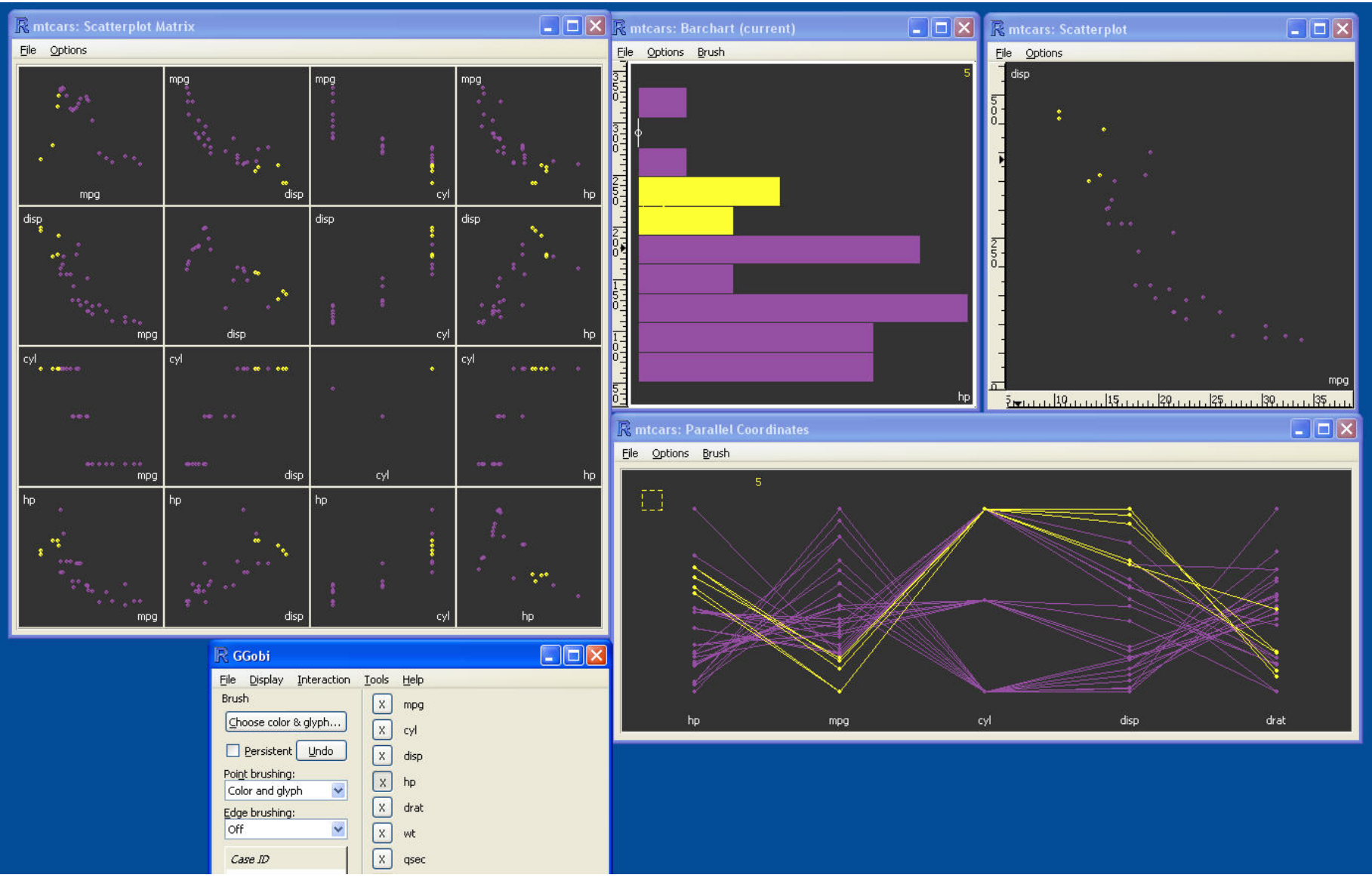
Why is that?

Two branches of interactive data visualization systems

- Started in 1970's
- Made by statisticians
- Data exploration
- Specialized but low code
- Target: researchers, working statisticians
- PRIM-9, XGobi/GGobi, Mondrian, iPlots

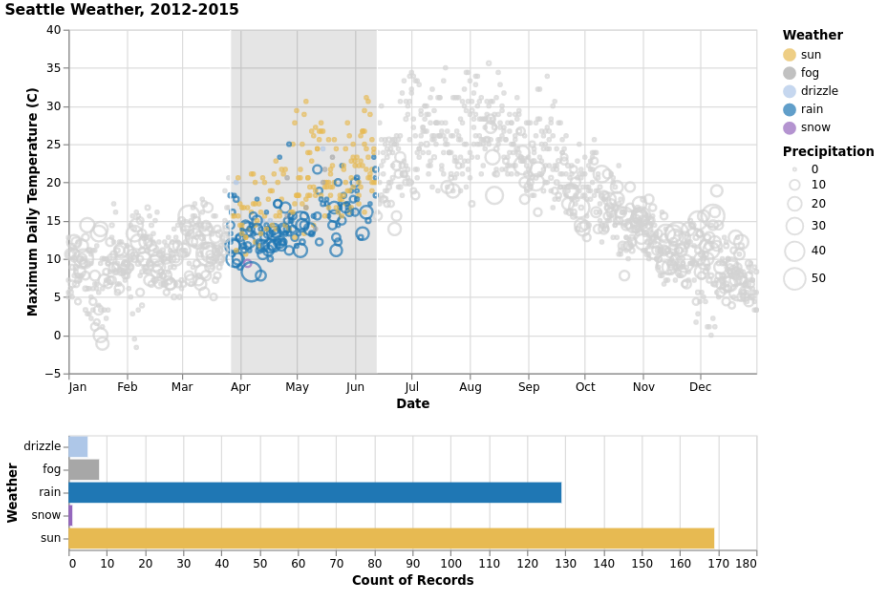
- Started in 2000's
- Made by CS's
- Data presentation
- Flexible but high code
- Target: anyone with a web browser
- D3js, highcharter, Vega




mtcars: Scatterplot Matrix
File Options
mpg
disp
cyl
hp
mtcars: Barchart (current)
File Options Brush
5
mtcars: Scatterplot
File Options
disp
mpg
mtcars: Parallel Coordinates
File Options Brush
5
hp
mpg
cyl
disp
drat
GGobi
File Display Interaction Tools Help
Brush
Choose color & glyph...
Persistent
Undo
Point brushing:
Color and glyph
Edge brushing:
Off
Case ID
mpg
cyl
disp
hp
drat
wt
qsec

Seattle Weather Exploration

This graph shows an interactive view of Seattle's weather, including maximum temperature, amount of precipitation, and type of weather. By clicking and dragging on the scatter plot, you can see the proportion of days in that range that have sun, rain, fog, snow, etc. Created by @jakevdp.



View this example in the online editor

Vega-Lite JSON Specification


```
{
  "$schema": "https://vega.github.io/schema/vega-lite/v5.json",
  "title": "Seattle Weather, 2012-2015",
  "data": {
    "url": "data/seattle-weather.csv"
  },
  "vconcat": [
    {
      "encoding": {
        "color": {
          "condition": {
            "param": "brush",
            "title": "Weather",
            "field": "weather",
            "type": "nominal",
            "scale": {
              "domain": ["sun", "fog", "drizzle", "rain", "snow"],
              "range": ["#e7ba52", "#a7a7a7", "#aec7e8", "#1f77b4", "#9467bd"]
            }
          },
          "value": "lightgray"
        },
        "size": {
          "title": "Precipitation",
          "field": "precipitation",
          "scale": {"domain": [-1, 50]},
          "type": "quantitative"
        },
        "x": {
          "field": "date",
          "timeUnit": "monthdate",
          "title": "Date",
          "axis": {"format": "%b"}
        },
        "y": {
          "title": "Maximum Daily Temperature (C)",
          "field": "temp_max",
          "scale": {"domain": [-5, 40]},
          "type": "quantitative"
        }
      },
      "width": 600,
      "height": 300,
      "mark": "point",
      "params": [{
        "name": "brush",
        "select": {"type": "interval", "encodings": ["x"]}
      }],
      "transform": [{"filter": {"param": "click"}}]
    },
    {
      "encoding": {
        "color": {
          "condition": {
            "param": "click",
            "field": "weather",
            "scale": {
              "domain": ["sun", "fog", "drizzle", "rain", "snow"],
              "range": ["#e7ba52", "#a7a7a7", "#aec7e8", "#1f77b4", "#9467bd"]
            }
          },
          "value": "lightgray"
        },
        "x": {"aggregate": "count"},
        "y": {"title": "Weather", "field": "weather"}
      },
      "width": 600,
      "mark": "bar",
      "params": [{
        "name": "click",
        "select": {"type": "point", "encodings": ["color"]}
      }],
      "transform": [{"filter": {"param": "brush"}}]
    }
  ]
}
```

Why this matters?

- Flexible, high-code systems are great for making interactive visualizations that many people will see and use

However:

- Time/effort cost may be too high for $n = 1$ audience
- Have to re-learn what you can do and expect
- Not everything that can be made should be made (footgun)

> "Interactivity appeared only once in our study, in a sketch; this indicates that the desire to build interactive views is present within the data science community, but the costs of using the tools outweigh the need during initial exploration."
>
> Batch & Elmquist, 2017

(sample size of 8, but is this surprising?)

Data exploration vs. presentation

- Two very different ecological niches
- Sometimes fundamentally different goals:

> "Graphical excellence is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space."
>
> *[...and least interaction?]*
>
> Tufte, 2001

Plotscaper: the idea

- Goal: an interactive data exploration tool that researchers and data scientists will actually use
- Aiming for the *Goldilocks zone*:
  - Low-code enough so that exploration is cost-effective
  - Flexible enough to allow for a range of visualizations
- Interactions consistent across different plot types
- Easily set up multi-plot figures with shared data
  - *What single-plot interactions are actually useful?*
- Follow data visualization principles, use sensible defaults

Plotscaper: where it is now

```
devtools::install_github("bartonicek/plotscaper")
```

- 5 wrapper plot types:
  - scatterplot, barplot, histogram, bubbleplot, squareplot
  - aiming for `ggplot2`-like aesthetic specification
- Currently supported interactivity:
  - transient/persistent selection
  - panning, changing size and alpha
- CSS grid layout for interesting figures
- Just a JavaScript file: Runs in RStudio viewever, RMarkdown, website, etc...

References

- Batch, Andrea, and Niklas Elmqvist. 2017. “The Interactive Visualization Gap in Initial Exploratory Data Analysis.” *IEEE Transactions on Visualization and Computer Graphics* 24 (1): 278–87.
- Tufte, Edward R. 2001. *The Visual Display of Quantitative Information*. Cheshire, Connecticut: Graphics Press LLC.