

Advanced Data Visualizations with **Stata**: Part II

Asjad Naqvi

International Institute for Applied Systems Analysis (IIASA)
Wirtschaftsuniversität Wien (WU)

Oceania Stata Conference
10 February 2022

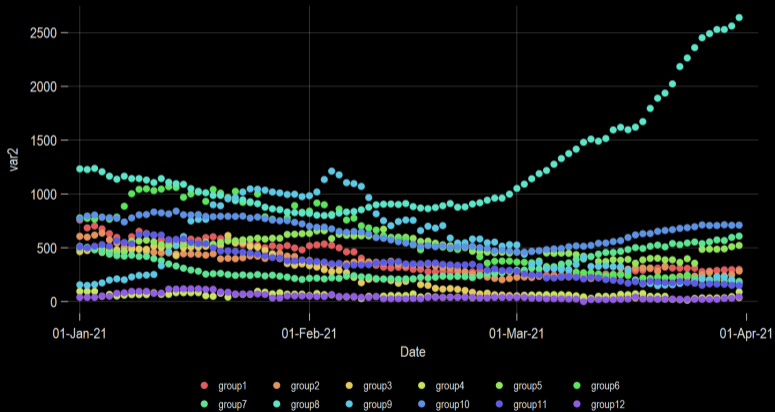
Background

- Currently based in Vienna, Austria
- Some 20 years of research + academia
- Started using Stata in 2003
- Have worked on numerous research projects using Stata
 - Data management, workflow, analysis
- Why do all the dataviz stuff?
 - Explosion of COVID-19 related dataviz + work-from-home + curiosity
 - Huge positive response from the online community

Why we need more #StataViz?

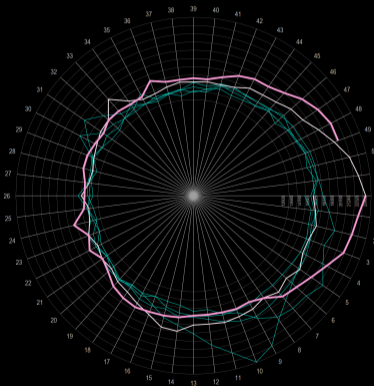
- Few official releases for graphs
- Still lots of options available in the default Stata structure
- Lots of great development in terms of dataviz packages:
 - `colorpalette`, `colrspace`, `heatplot` (Jann), `spmap` (Pisati), `geo2xy`, `geonear`, `geoinpoly` (Piccard)
- What I will discuss here:
 - Part1: Polar plots
 - Part2: Maps
 - Overview of data visualizations

Stata schemepack

Scatter plot
Scheme neon

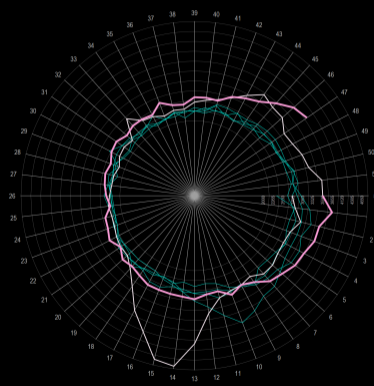
Polar plots

Excess weekly deaths for 65+ - DE



Source: Eurostat table demo_z_mweeK3, 2020 = light pink, 2021 = dark pink, and 2019-2019 = green shades.

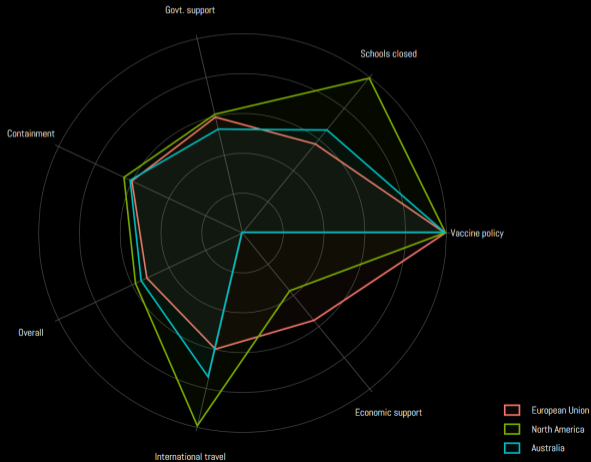
Excess weekly deaths for 65+ - NL



Source: Eurostat table demo_z_mweeK3, 2020 = light pink, 2021 = dark pink, and 2019-2019 = green shades.

Spider plots

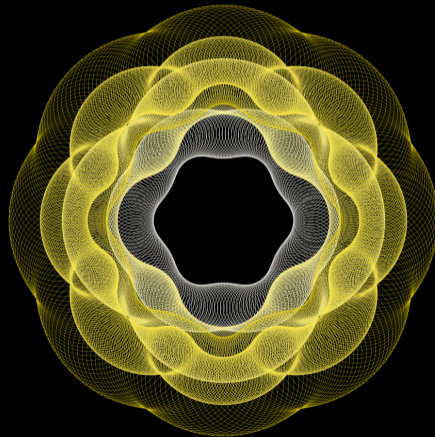
COVID-19 policies: 04 Feb 2022



Data source: Oxford COVID-19 Government Response Tracker.

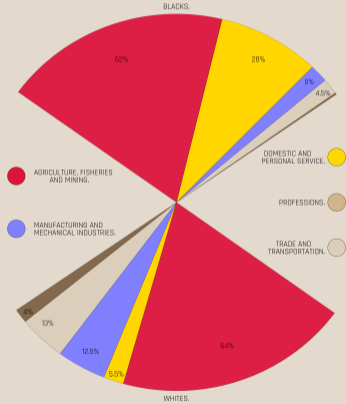
Isaac Asimov, *The Robot Mind* (1983)

I believe in evidence, confirmed by independent observers. I'll believe anything, however, the firmer and more solid the evidence will have to be. The wilder and more ridiculous something is, no matter how wild and ridiculous, if there is evidence for it. I believe in observation, measurement, and reasoning, and I'll believe in evidence, confirmed by independent observers. I'll believe anything, however, the firmer and more solid the evidence will have to be.



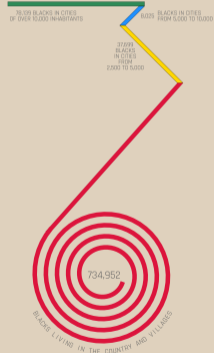
Du Bois

OCCUPATION OF BLACKS AND WHITES IN GEORGIA.



#DuBoisChallenge Nr. 3 made with #Stata. Source: <https://github.com/mjstankis/dubois-data-panels>, Plate 27. By Agad Nazki (agadnazki@gmail.com).

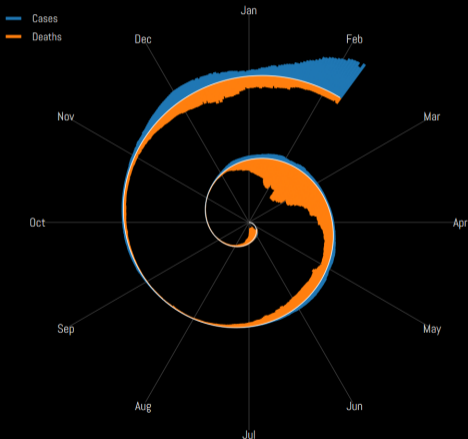
CITY AND RURAL POPULATION. 1890.



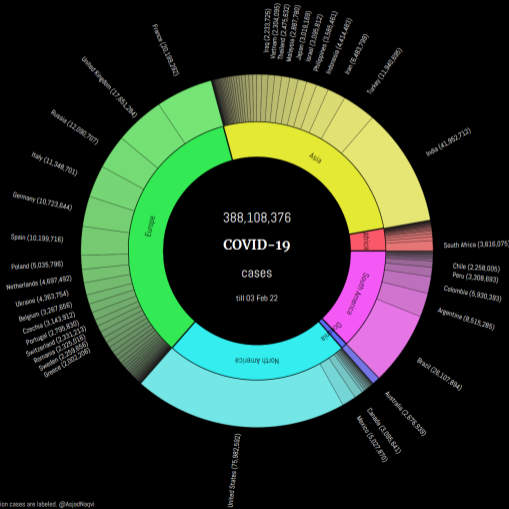
#DuBoisChallenge Nr. 8 made with #Stata. Source: <https://github.com/mjstankis/dubois-data-panels>, Plate 8. By Agad Nazki (agadnazki@gmail.com).

Snakeplot

Daily Cases and Deaths in Germany (2020-2022)

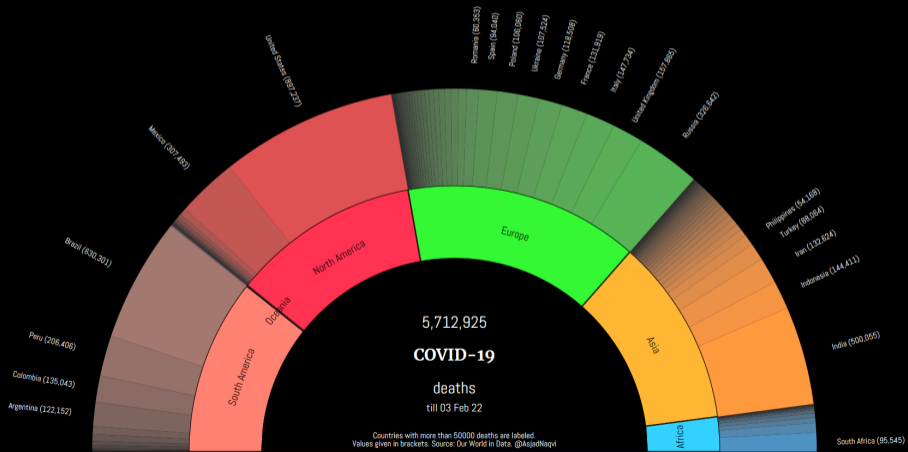


Sunburst plot

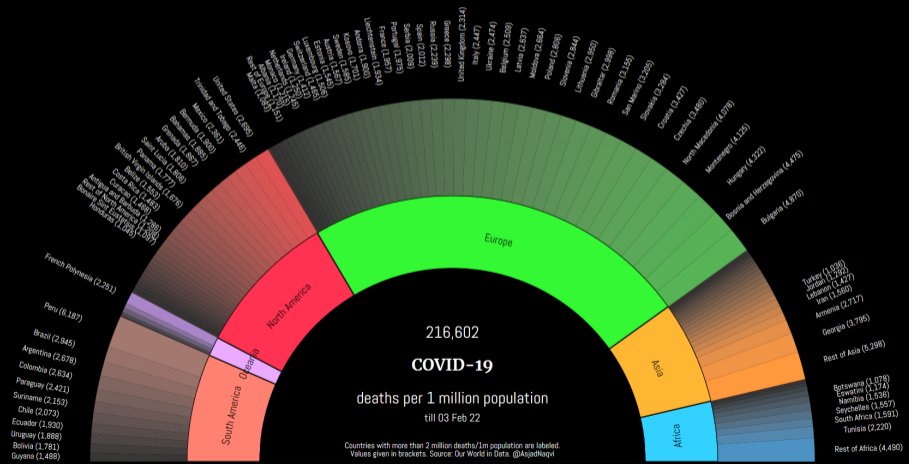


Source: Our World in Data. Countries with >2 million cases are labeled. @AqadNazi

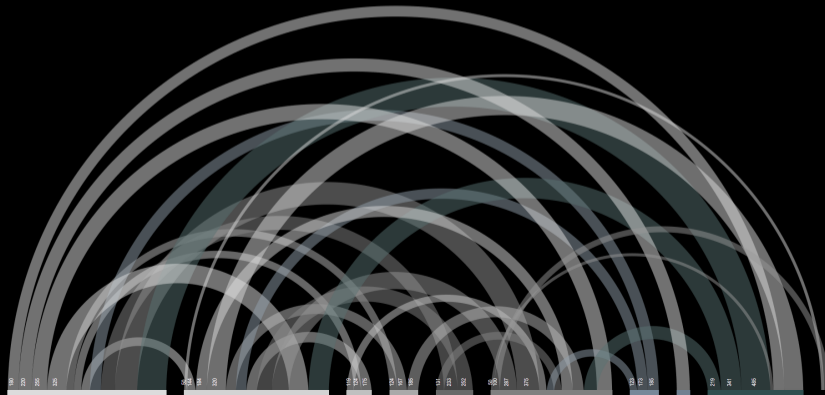
Half sunburst plot



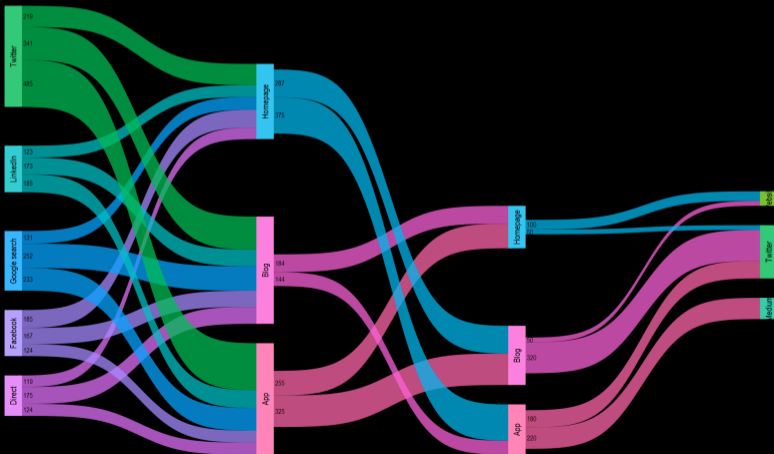
Half sunburst plot



Arc plot

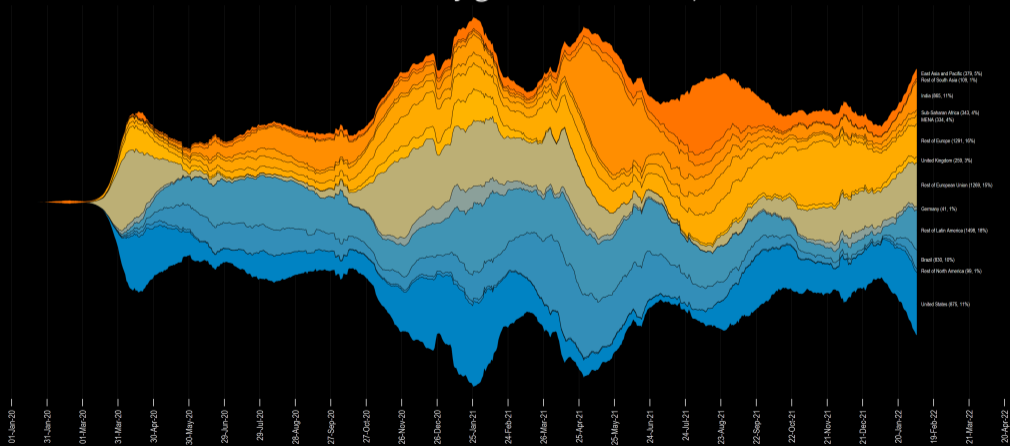


Sankey



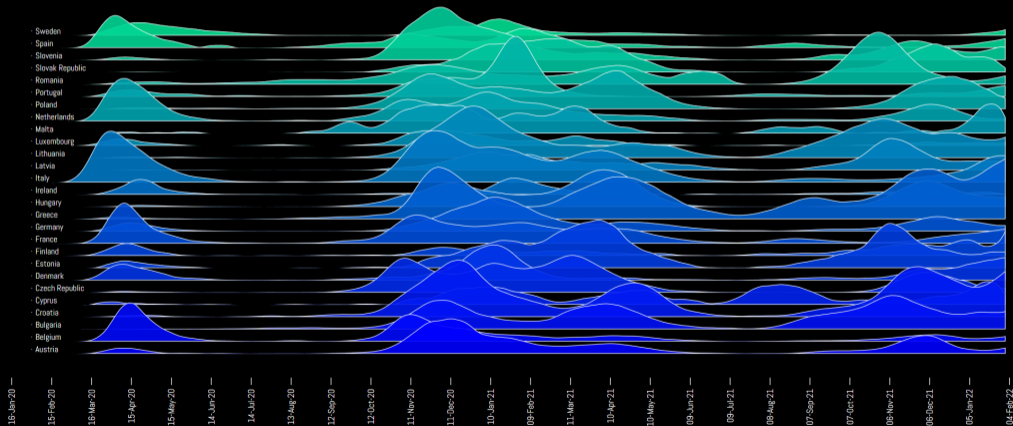
Stream plots

COVID-19 daily global deaths: 8,292



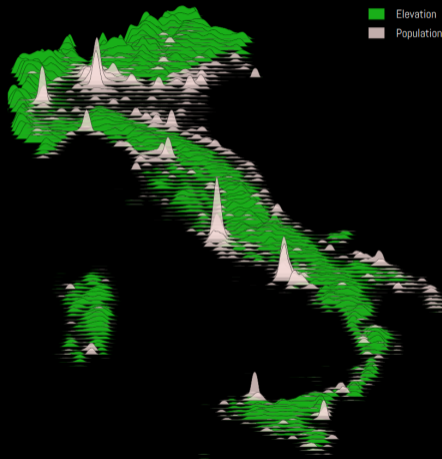
Joyplots

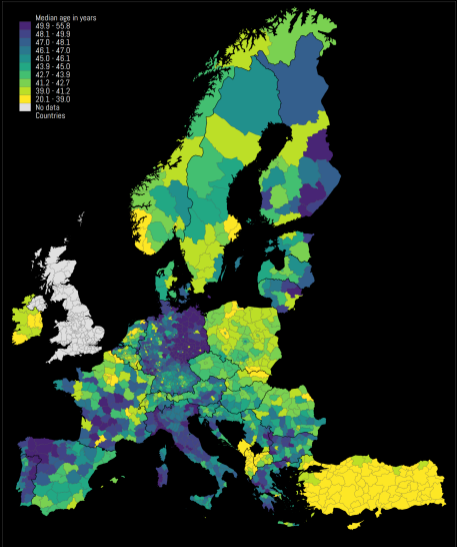
COVID-19 daily deaths in the EU

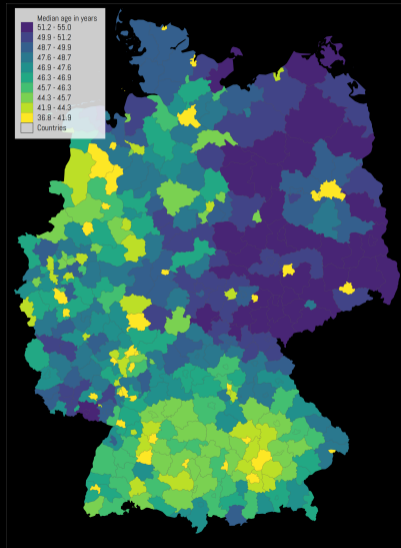


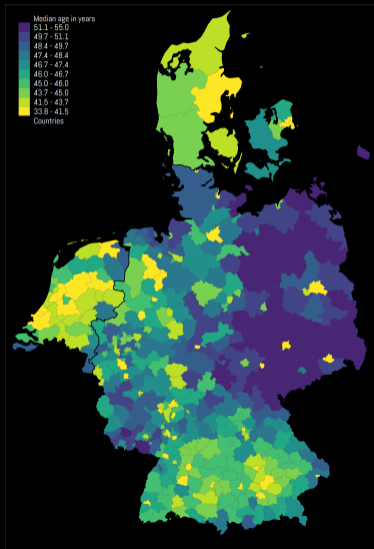
Data sources: Our World in Data, World Bank classifications used for country groups. Each country plot is normalized by its maximum value.

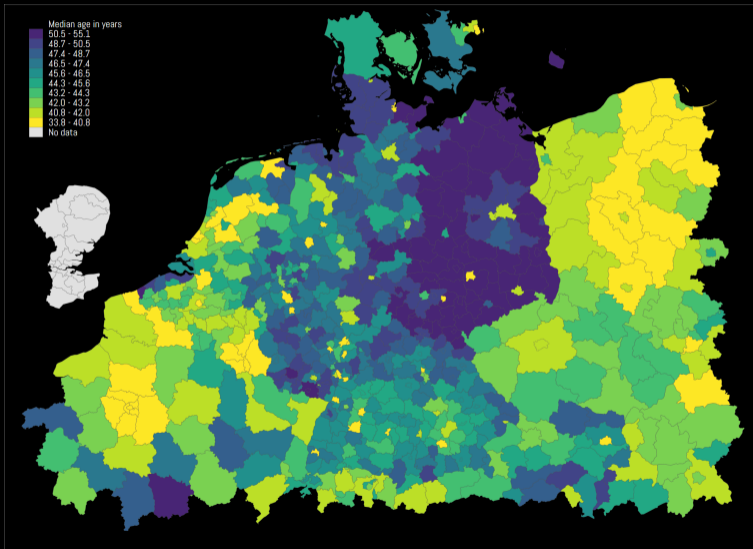
Italy: Population vs Elevation

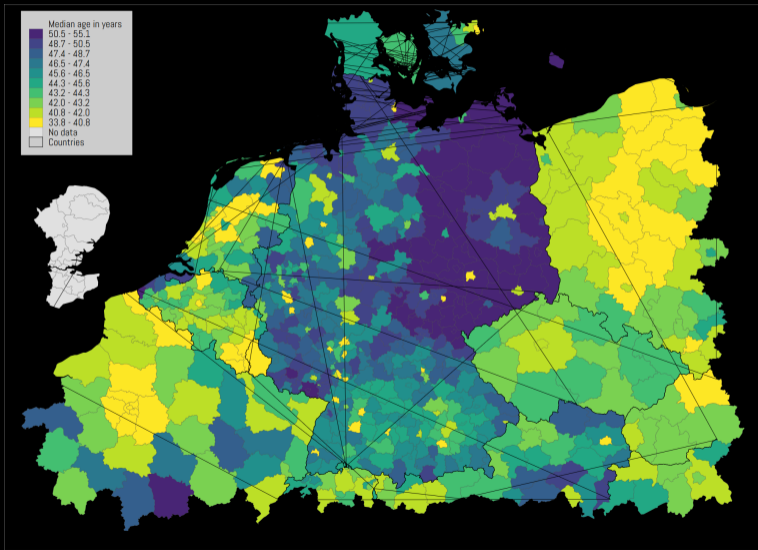






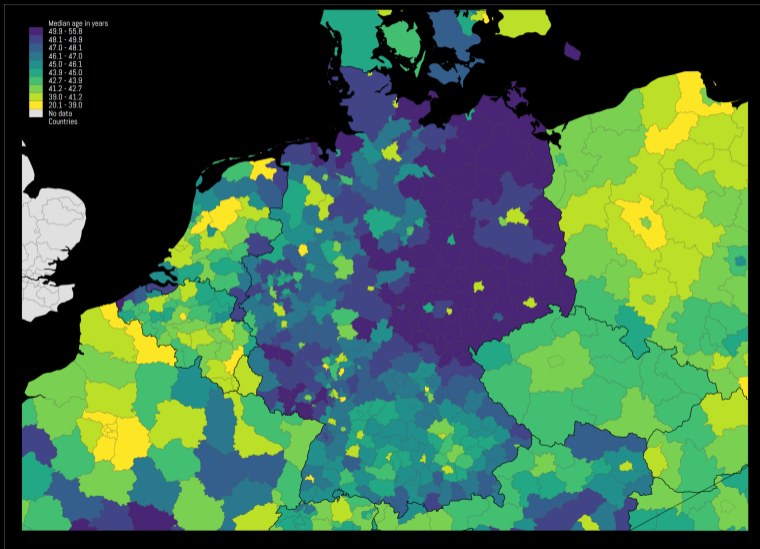




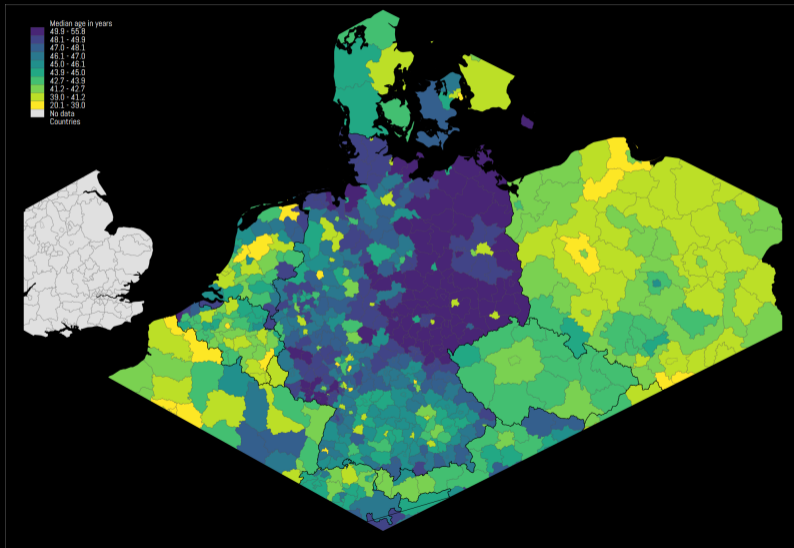


Sutherland–Hodgman polygon clipping algorithm

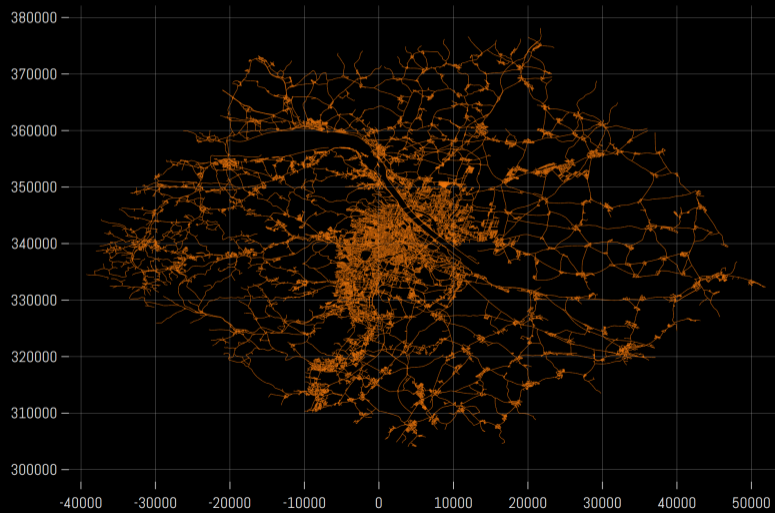
- Input 1: Polygon _shp file: Should be sorted clockwise or counter-clockwise
- Input 2: Bounding convex shape: Needs to be **convex** and sorted **clockwise** and fully closed
- Stata psuedo script:
 - 1 Fully identify all the “islands” in *_shp (Additional step post spshape2dta)
 - 2 Identify points in *_shp file which are inside and outside the bounds (geoinpoly by Picard 2015)
 - 3 Throw out the polygons that are fully outside
 - 4 Separate fully inside and partially inside
 - 5 Process only the partially inside using the S-H algorithm
 - 6 Ensure shape order (a lot of reshaping here)
 - 7 Save as clipped shapefile



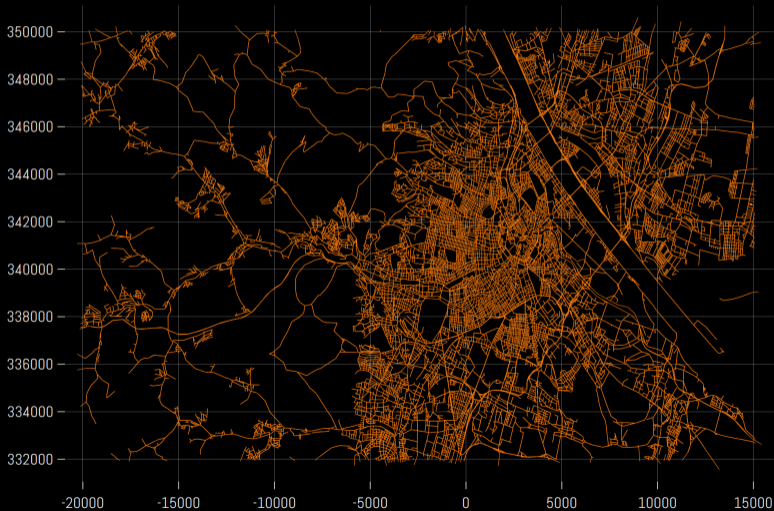
Hexagon clipping



Line clipping (full extent)

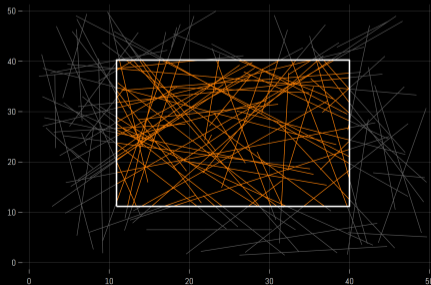


Line clipping (using a box)

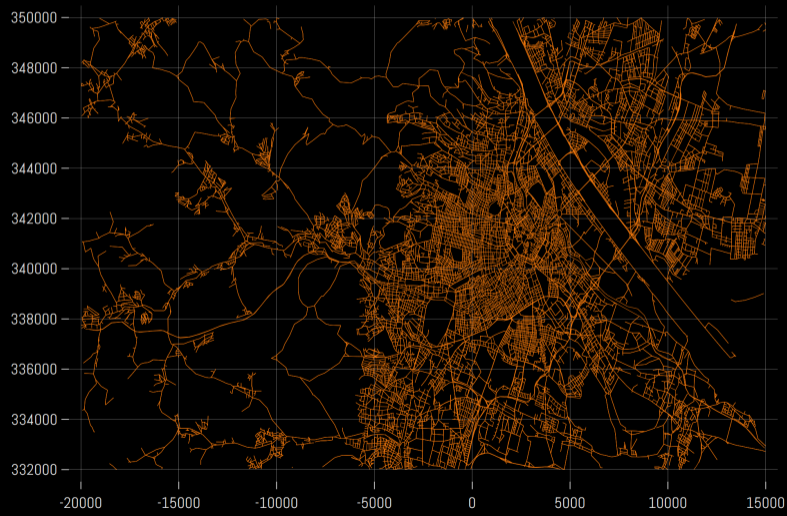


Cohen–Sutherland line clipping algorithm

- Input 1: Line endpoints (2 coordinate pairs)
- Input 2: Bounding box
- Stata psuedo script:
 - 1 Mark lines that are fully outside and inside
 - 2 Partially inside lines are clipped on the box intersection
- Beta version on GitHub: [clipline](#)



Line clipping (clipping algorithm)



Line clipping (spmap)



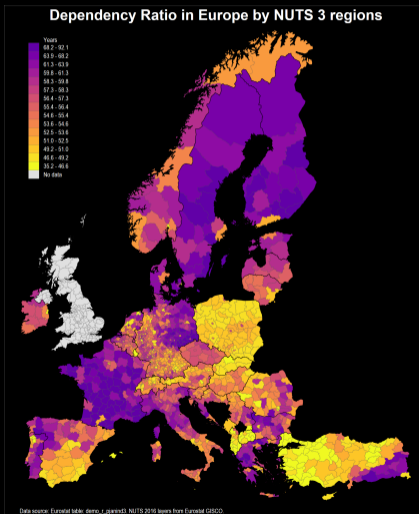
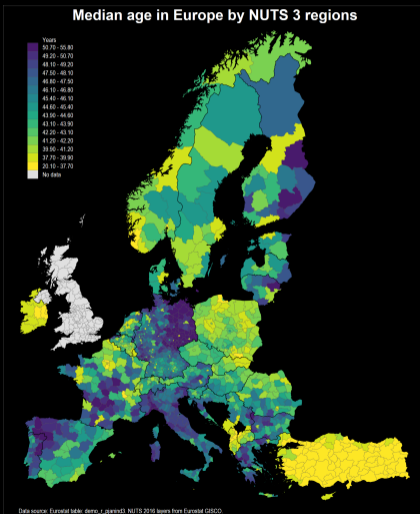
Line clipping (spmap)



Scatter plot extracted from rasters



Maps II guide



custom map + spikes

Social Connectivity Index - UK

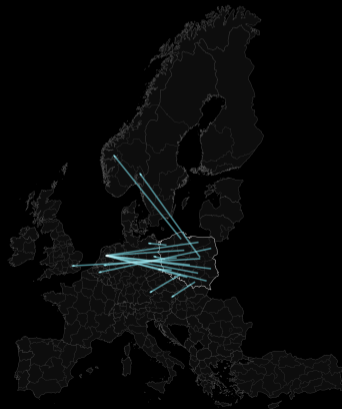
Lines show NUTS 2 regions with the highest connectivity outside the origin country



Map layers: Eurostat, SD data Facebook

Social Connectivity Index - Poland

Lines show NUTS 2 regions with the highest connectivity outside the origin country

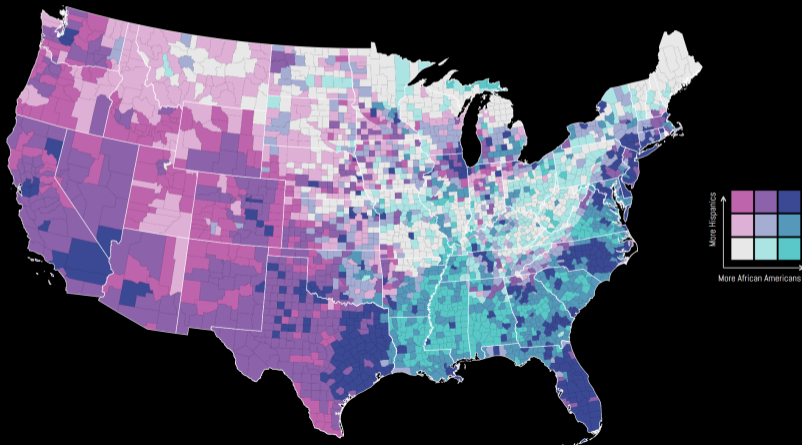


Map layers: Eurostat, SD data Facebook

Bi-variate map

The USA Race Space

County-level share of African Americans vs Hispanics



Source of data, and map layers: United States Census Bureau. 2020 estimates. Tercile cut-offs of shares used for color categories.

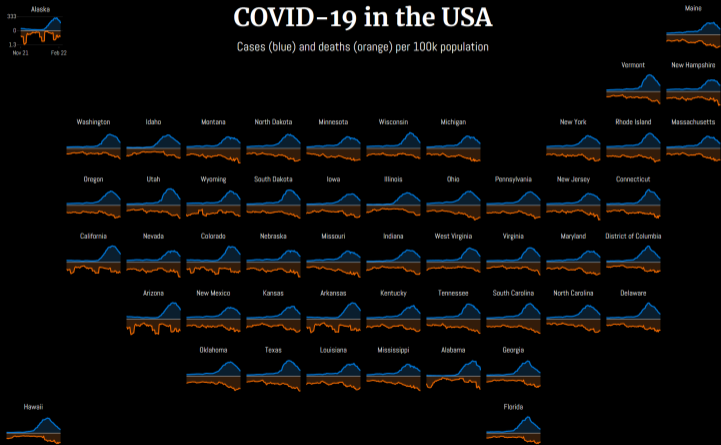
custom map + QGIS



Tilemap

COVID-19 in the USA

Cases (blue) and deaths (orange) per 100k population



Source: New York Times Github repository (<https://github.com/nytimes/covid-19-data>). 7-day moving average used to smooth the data. @MojibNajafi

Tilemap

COVID-19 in the USA

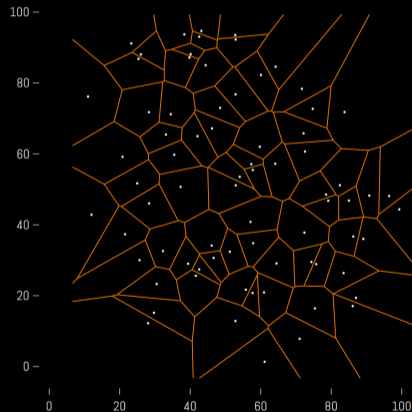
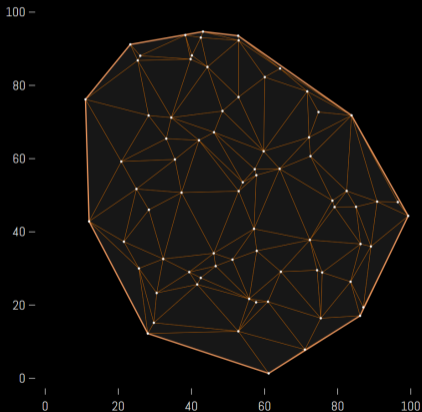
Cases (blue) and deaths (orange) per 100k population



Source: New York Times GitHub repository (<https://github.com/nytimes/covid-19-data>). 7-day moving average used to smooth the data. @MojibNajafi

Delaunay triangles + Voronoi tessellations + Convex hull

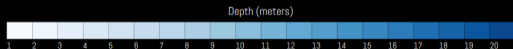
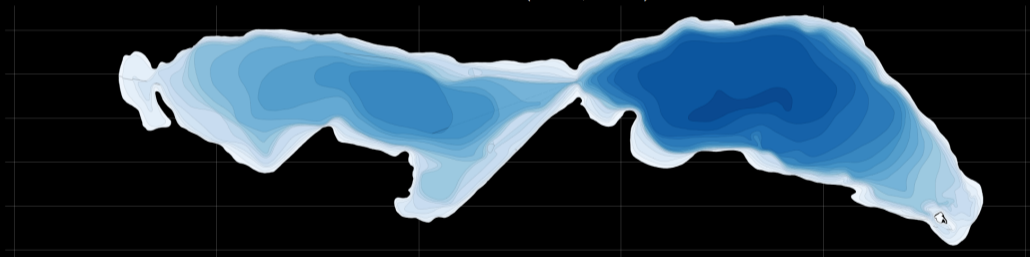
Beta package on GitHub: [delaunay](#)



Bezier curves and splines

Bathymetric contours




Lesser Slave Lake (Alberta, Canada)



Map layer: Open data provided by the Government of Alberta, Canada.

More info




More Stata viz:

-  [The Stata Guide on Medium](#)
-  [#30DayMapChallenge 2021](#)
-  [The Stata Gallery](#)

Activities:

- Recently featured on the World Bank's [Development Impact](#) blog
- Regular online courses on dataviz with Stata

Connect with me:

- @ asjadnaqvi@gmail.com
-  github.com/asjadnaqvi
-  [@AsjadNaqvi](#)
-  [AsjadNaqvi](#)

Wish list!

- Ability to control marker size scaling
- Fully flexible marker clock position
- Line weights (like marker weights)
- Allow angles/sizes/colors to be read from variables (like mlabels)
- Color/alpha scaling for markers/lines
- Color gradients for lines/areas
- Ability to add custom markers (or increase marker pool)
- Ability to added colored text in graphs
- Ability to add images in figures
- Better graph combine options
- Ability to read images (pixel data/colors)
- Ability to draw/add/subtract areas between two or more functions