Advanced Data Visualizations with Stata: Part VI

Visualizing three variables

Asjad Naqvi

Austrian Institute of Economic Research (WIFO)

Stata Northern European Conference – Oslo, Norway
10 September 2024

StataViz

- Lots of exciting developments in terms of dataviz packages in the past years.
- Base colors and custom color palettes use colorpalette, colrspace by Ben Jann.
- Own packages (21 currently): alluvial, arcplot, bimap, bumparea, bumpline, circlebar, circlepack, clipgeo, delaunay, joyplot, marimekko, polarspike, sankey, schemepack, spider, splinefit, streamplot, sunburst, treecluster, treemap, waffle
- Today we will discuss two new packages that deal with plotting three variables.

- Any three variables can be plotted in a 3-D space.

- Any three variables can be plotted in a 3-D space.
- But if three variables add up to a 1 or a 100, we can use Barycentric projection and plot them in a triangle. Some examples:
  - Shares of population in different age groups: 0-14, 15-64, 65 and above.
  - Shares of primary (agriculture), secondary (manufacturing), and tertiary (services) sectors.
  - Shares of primary, secondary, tertiary education attainment level in working-age populations.

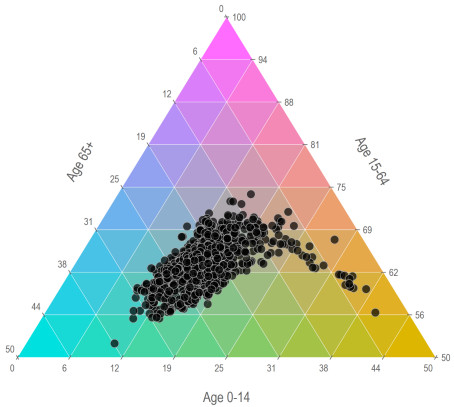
- Any three variables can be plotted in a 3-D space.
- But if three variables add up to a 1 or a 100, we can use Barycentric projection and plot them in a triangle. Some examples:
  - Shares of population in different age groups: 0-14, 15-64, 65 and above.
  - Shares of primary (agriculture), secondary (manufacturing), and tertiary (services) sectors.
  - Shares of primary, secondary, tertiary education attainment level in working-age populations.
- triplot by Nick Cox (2009) already exists that can visualize three variables.
  - Not very customizable and limited colors.
  - Allows for data recentering (can be confusing to interpret).

- Any three variables can be plotted in a 3-D space.
- But if three variables add up to a 1 or a 100, we can use Barycentric projection and plot them in a triangle. Some examples:
  - Shares of population in different age groups: 0-14, 15-64, 65 and above.
  - Shares of primary (agriculture), secondary (manufacturing), and tertiary (services) sectors.
  - Shares of primary, secondary, tertiary education attainment level in working-age populations.
- triplot by Nick Cox (2009) already exists that can visualize three variables.
  - Not very customizable and limited colors.
  - Allows for data recentering (can be confusing to interpret).
- Introducing ternary, a command with an easy-to-use syntax and multiple color options with a complete color engine in the backend.

ternary

A Stata package for tri-variate plots

ssc install ternary, replace
GitHub: https://github.com/asjadnaqvi/stata-ternary



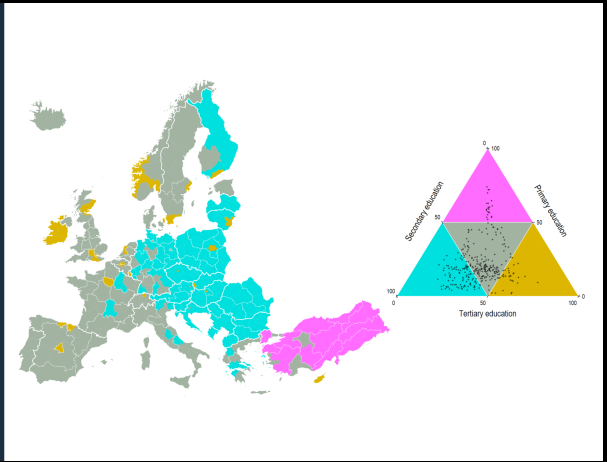
Let's check it out!

- We can also extend ternary to maps
- This allows us to explore spatial distribution of three variables and their relationships using the trimap command.
  - Similar to bimap.
  - Uses the full funcationality of ternary and geoplot.
  - Fully customize maps and colors.

trimap

A Stata package for tri-variate maps

ssc install trimap, replace
GitHub: https://github.com/asjadnaqvi/stata-trimap



Let's check it out!

What can be improved?

- Stata can link with GEOS and/or GDAL libraries to add full functionality of the maps.
  - These already form the base engine for most R and Python packages and used in dedicated GIS softwares such as ArcGIS and QGIS.
- Allow angles, sizes, colors, line markers, clock positions, etc. to be read from variables or macro lists (like mlabels).
- Modifiable text colors.
- Better or more custom markers.
- Figure-in-figure capability.

Thank you!

More Stata viz:

- The Stata Guide on Medium
- The Stata Gallery on Medium
- StataViz portfolio
- #30DayMapChallenge 2021

Connect with me:

- asjadnaqvi@gmail.com
- github.com/asjadnaqvi
- @AsjadNaqvi
- AsjadNaqvi