Introduction
○○○

Stocks
○○○○○○○○○○○○○○○○○○○

Flows
○○○○○

Last slides
○○

Advanced Data Visualizations with Stata: Part V
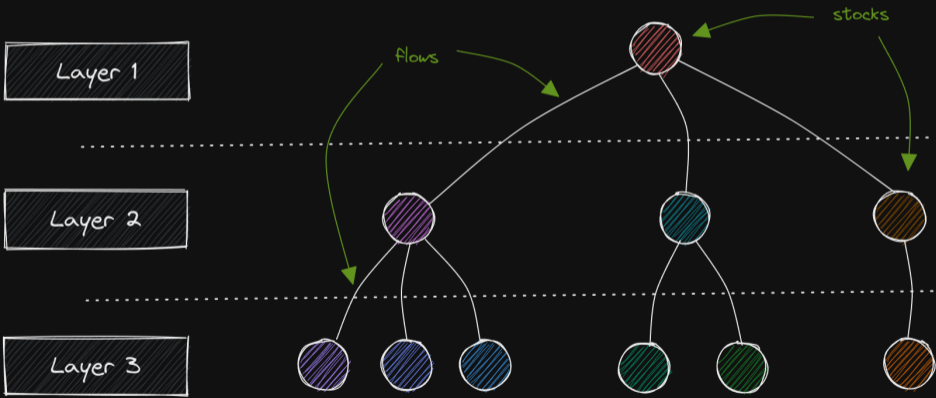
Hierarchical data

Asjad Naqvi

Austrian Institute of Economic Research (WIFO)
Vienna University of Economics and Business (WU)
International Institute for Applied Systems Analysis (IIASA)

Stata Oceania Conference
9 February 2023

Introduction
●○○

Stocks
○○○○○○○○○○○○○○○○○○○

Flows
○○○○○

Last slides
○○

Why we need more #StataViz?

- A lot of potential given the existing Stata/Mata structure
- Lots of exciting developments in terms of dataviz packages in the past years.
- Own packages: bimap, joyplot, streamplot, clipgeo, delaunay, marimekko, arcplot, treemap, circlepack, circlebar, spider, sankey, alluvial, treecluster
- Plus several other packages by the community have been released

- Today we will discuss some of these packages that deal with heirarchical data

Introduction
○●○

Stocks
○○○○○○○○○○○○○○○○

Flows
○○○○○

Last slides
○○

Hierarchical data

Layer 1

Layer 2

Layer 3

flows

stocks

Introduction
○○●

Stocks
○○○○○○○○○○○○○○○○○○○

Flows
○○○○○

Last slides
○○

- Data visualization dealing with hierarchical data can be split into three groups
  - Stocks
  - Flows
  - Stocks and flows
- Adding other dimensions, like time, makes the data structures fairly complex
- Existing packages make these visualizations easier:
  - circlepack, circlebar, treemap, sankey, treecluster
- We will discuss some applications using the same data

Introduction
○○○

Stocks
●○○○○○○○○○○○○○○○○○

Flows
○○○○○

Last slides
○○

Hierarchical data examples

- Hospitals, doctors, patients
- Schools, teachers, students
- Budgets
- Populations at different admin units

228	DE	DE2	DE22	DE221	73,411				
229	DE	DE2	DE22	DE222	52,803				
230	DE	DE2	DE22	DE223	47,791				
231	DE	DE2	DE22	DE224	119,478				
232	DE	DE2	DE22	DE225	78,362				
233	DE	DE2	DE22	DE226	123,058				
234	DE	DE2	DE22	DE227	159,895				
235	DE	DE2	DE22	DE228	192,656				
236	DE	DE2	DE22	DE229	77,410				
237	DE	DE2	DE22	DE22A	121,502				
238	DE	DE2	DE22	DE22B	101,120				
239	DE	DE2	DE22	DE22C	96,683				
240	DE	DE2	DE23	DE231	42,207				
241	DE	DE2	DE23	DE232	153,094				

Introduction
○○○

Stocks
○●○○○○○○○○○○○○○○○○

Flows
○○○○○

Last slides
○○

Single layer stocks

Stocks of single layers can be easily visualized using in-built packages, for example population of Spanish NUTS1 regions

Introduction
○○○

Stocks
○○●○○○○○○○○○○○○○○○

Flows
○○○○○

Last slides
○○
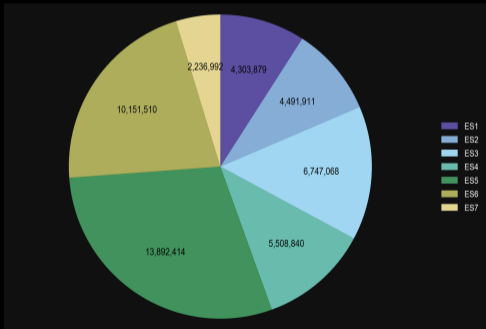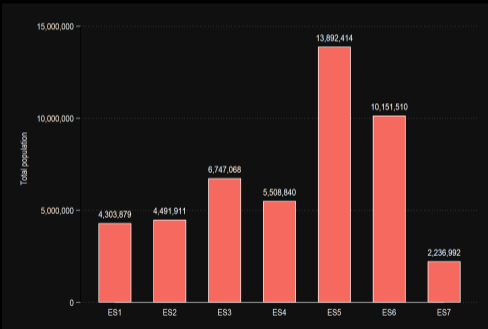
Single layer packages

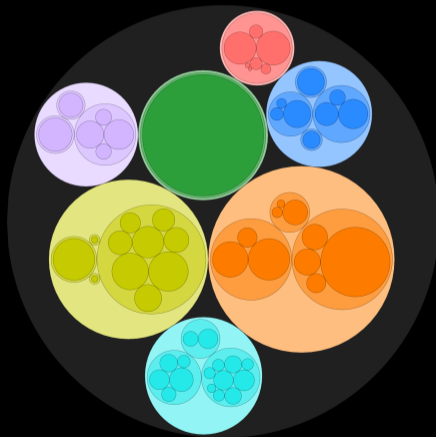Stocks of single layers can also be easily visualized using custom packages, e.g. circlebar

Introduction
○○○

Stocks
○○○●○○○○○○○○○○○○○○○

Flows
○○○○○

Last slides
○○

Hierarchical stocks: `circlepack`

- Plot nested lists as circles within circles.
- Circle packing requires a packing algorithm that maximize the area of a starting circle, such that no circles overlap.

$\rightarrow$ `ssc install circlepack, replace`

Introduction
○○○

Stocks
○○○○●○○○○○○○○○○○○○○

Flows
○○○○○

Last slides
○○

Hierarchical stocks: `circlepack`

Introduction
ooo

Stocks
oooooo●ooooooooooooo

Flows
ooooo

Last slides
oo

Hierarchical stocks: `treemap`

- Treemaps plot nested lists as rectangles within rectangles.
- treemap uses the "squarify" tiling algorithm (also used by OEC, OWID, D3js).
- "squarify" optimizes placement of tiles to approach a desired aspect ratio.
- Tiles are sorted by largest-to-smallest size, and iteratively placed by either "slicing" or "dicing" a starting rectangle, until all tiles are close to the desired aspect ratio.

$\rightarrow$ `ssc install treemap, replace`

Introduction
○○○

Stocks
○○○○○○●○○○○○○○○○○○

Flows
○○○○○

Last slides
○○

Hierarchical stocks: `treemap`

ES6 (10,151,510)

ES61
8,478,083

ES62
1,504,869

Rest of ES?
168,566

ES2 (4,491,911)

ES24
1,330,333

ES23
315,651

ES22
656,589

ES21
2,189,138

ES1 (4,303,879)

ES12
1,018,899

ES13
582,388

ES11
2,702,592

ES7 (2,236,992)

ES70
2,236,992

ES5 (13,892,414)

ES51
7,652,348

ES53
1,210,725

ES52
5,029,341

ES3 (6,747,068)

ES30
6,747,068

ES4 (5,508,840)

ES42
2,045,554

ES43
1,061,979

ES41
2,401,307

Introduction
○○○

Stocks
○○○○○○○○●○○○○○○○○○○○

Flows
○○○○○

Last slides
○○

Hierarchical stocks: `treemap`

Introduction
○○○

Stocks
○○○○○○○○●○○○○○○○○○○○

Flows
○○○○○

Last slides
○○

Hierarchical stocks: `sunburst`

- Layers are stacked on top of each other in a semi-circle.
- Arc lengths represent proportion
- Allows complete customization of layers

→ `ssc install sunburst, replace`

Introduction
○○○

Stocks
○○○○○○○○○●○○○○○○○○

Flows
○○○○○

Last slides
○○

Hierarchical stocks: `sunburst`

Introduction
○○○

Stocks
○○○○○○○○○●○○○○○○○

Flows
○○○○○

Last slides
○○

Hierarchical stocks: `sunburst`

Introduction
○○○

Stocks
○○○○○○○○○○○○●○○○○○○

Flows
○○○○○

Last slides
○○
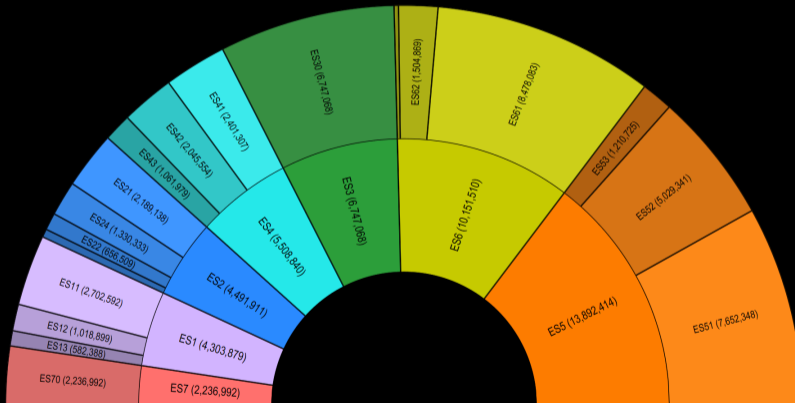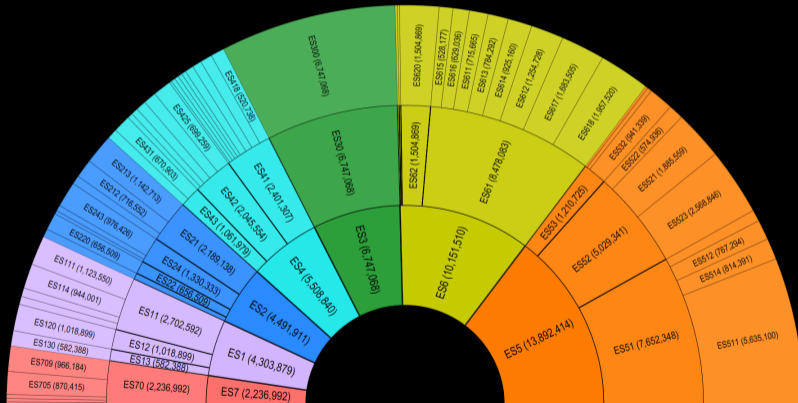
Hierarchical stocks: `sankey`

- Sub-division of stocks are shown as flows.
- Flows across categories.
- Allows complete customization of layers

→ `ssc install sankey, replace`

Introduction
○○○

Stocks
○○○○○○○○○○●○○○○○

Flows
○○○○○

Last slides
○○

Hierarchical stocks: `sankey`

Introduction
○○○

Stocks
○○○○○○○○○○○○●○○○○○

Flows
○○○○○

Last slides
○○

Hierarchical stocks: `sankey`

Introduction
○○○

Stocks
○○○○○○○○○○○○○○●○○○

Flows
○○○○○

Last slides
○○
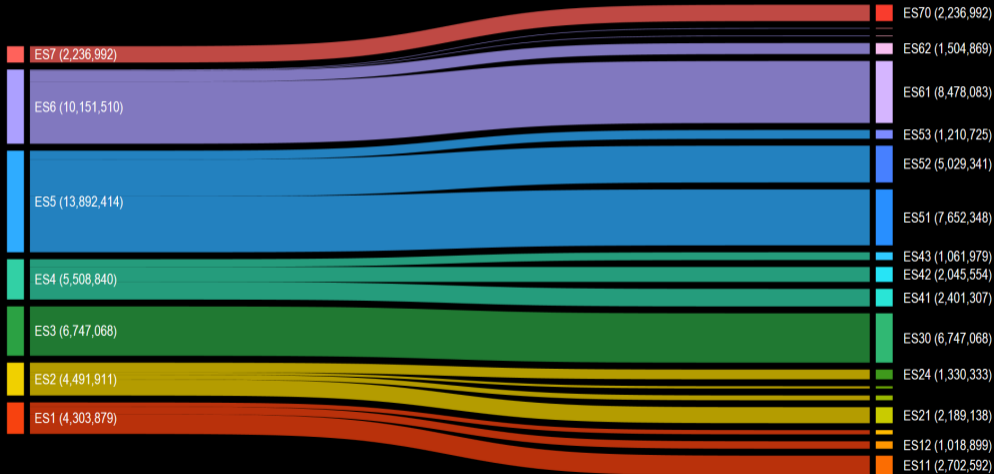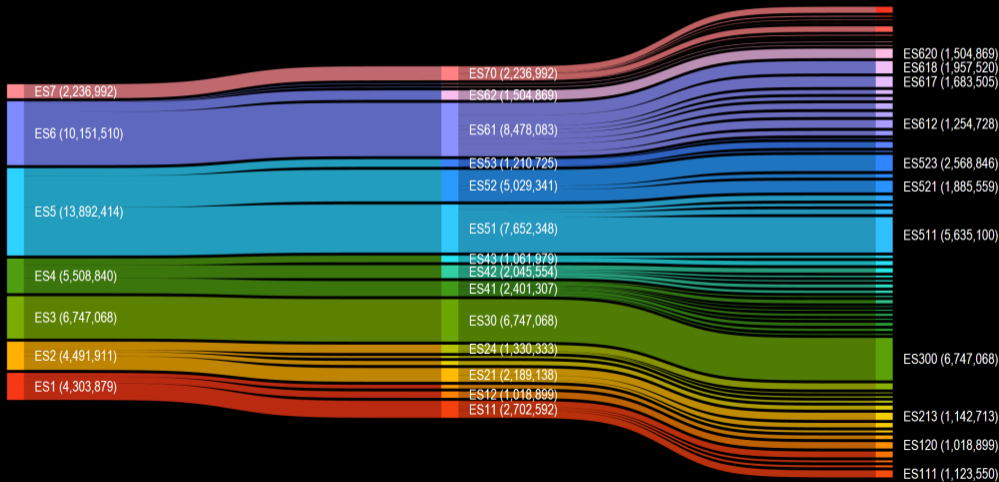
Hierarchical stocks: `treecluster`

- Layers are shown in a tree structure.
- Trees can be horizontal or radial
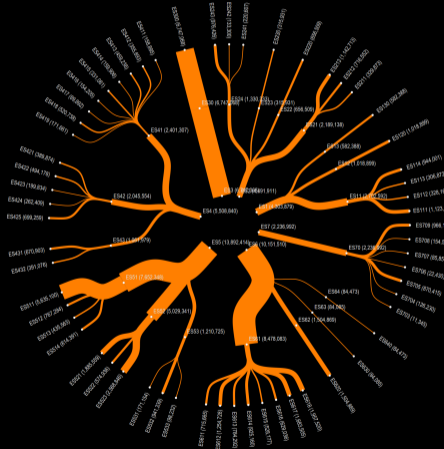- Allows complete customization of layers

$\rightarrow$ `ssc install treecluster, replace`
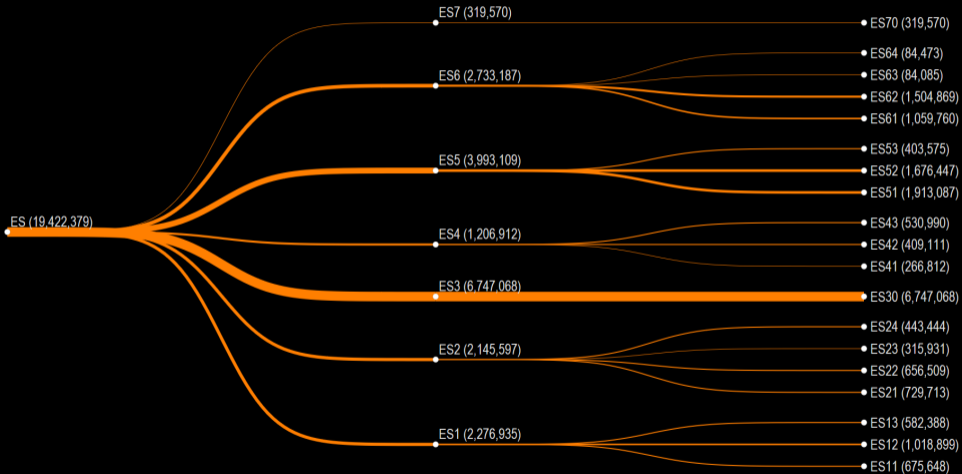
Introduction
○○○

Stocks
○○○○○○○○○○○○○●○○

Flows
○○○○○

Last slides
○○

Hierarchical stocks: `treecluster`

Introduction
○○○

Stocks
○○○○○○○○○○○○○●○

Flows
○○○○○

Last slides
○○

Hierarchical stocks: `treecluster`

Introduction
○○○

Stocks
○○○○○○○○○○○○○○●

Flows
○○○○○

Last slides
○○

Hierarchical stocks: `treecluster`

Introduction
ooo

Stocks
ooooooooooooooooooo

Flows
●oooo

Last slides
oo
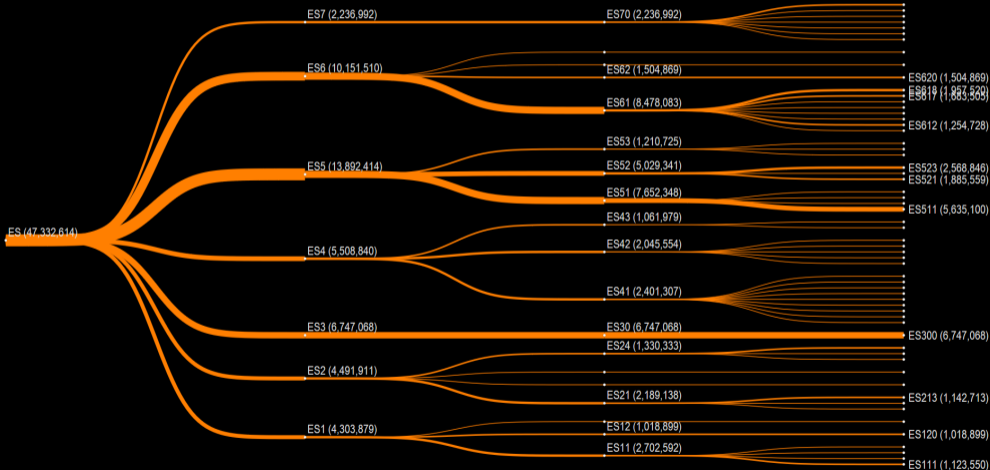
Hierarchical flows

- Hierarchical flows occur across and within layers.
- Even within layers bi-directional flows make data visualizations complex.
- We use the recently-published bi-lateral migration flows dataset:

  Abel, G., Cohen, J. (2022). Bilateral international migration flow estimates updated and refined by sex. Scientific Data 9, 173. DOI: https://doi.org/10.1038/s41597-022-01271-z

- The data is estimated for country-country pairs, at 5 year intervals, and by gender
- Data is collapsed at the sub-region, and all migration

Introduction
○○○

Stocks
○○○○○○○○○○○○○○○

Flows
○●○○○

Last slides
○○

Hierarchical flows: `heatplot`

Introduction
○○○

Stocks
○○○○○○○○○○○○○○○

Flows
○○●○○

Last slides
○○

Hierarchical flows: `arcplot`

Introduction
○○○

Stocks
○○○○○○○○○○○○○○○○○

Flows
○○○○●○

Last slides
○○

Hierarchical flows: `sankey`

Introduction
○○○

Stocks
○○○○○○○○○○○○○○○○

Flows
○○○○●

Last slides
○○

Hierarchical flows: chord

Introduction
○○○

Stocks
○○○○○○○○○○○○○○○○○○○

Flows
○○○○○

Last slides
●○
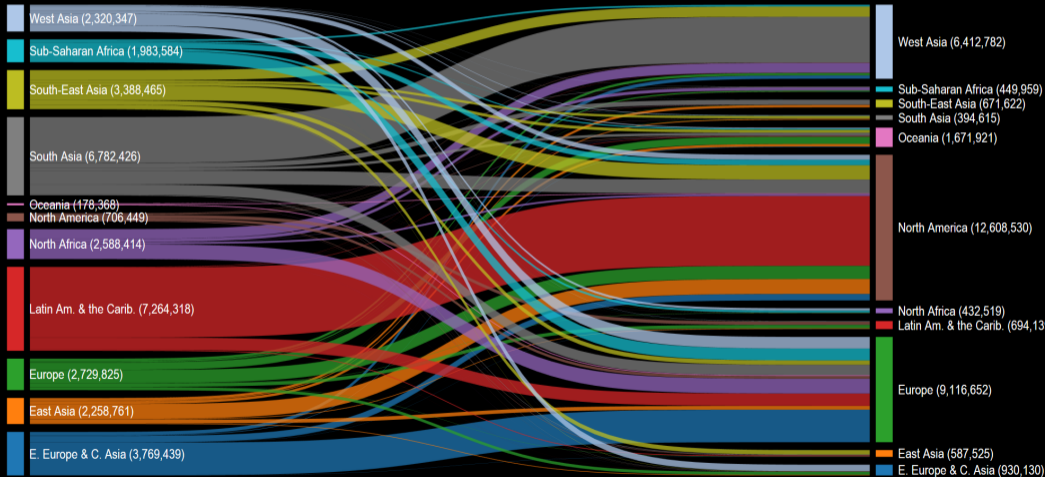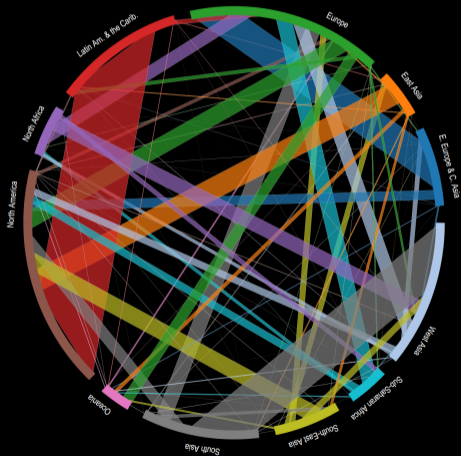
What can be improved

- **Ability to read image data**
- Plotting by colors
- Making shapes, for example, circles should be easier
- Allow angles, sizes, colors, clock positions, etc. to be read from variables or macro lists (like mlabels)
- Line weights (like marker weights)
- Allow adding custom markers (or increase marker pool)
- Allow colored text in graphs
- Ability to overlay images in figures

Introduction
○○○

Stocks
○○○○○○○○○○○○○○○○○○

Flows
○○○○○

Last slides
○●

More info

More Stata viz:

- The Stata Guide on Medium
- The Stata Gallery on Medium
- #30DayMapChallenge 2021
- #30DayMapChallenge 2022

Connect with me:

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