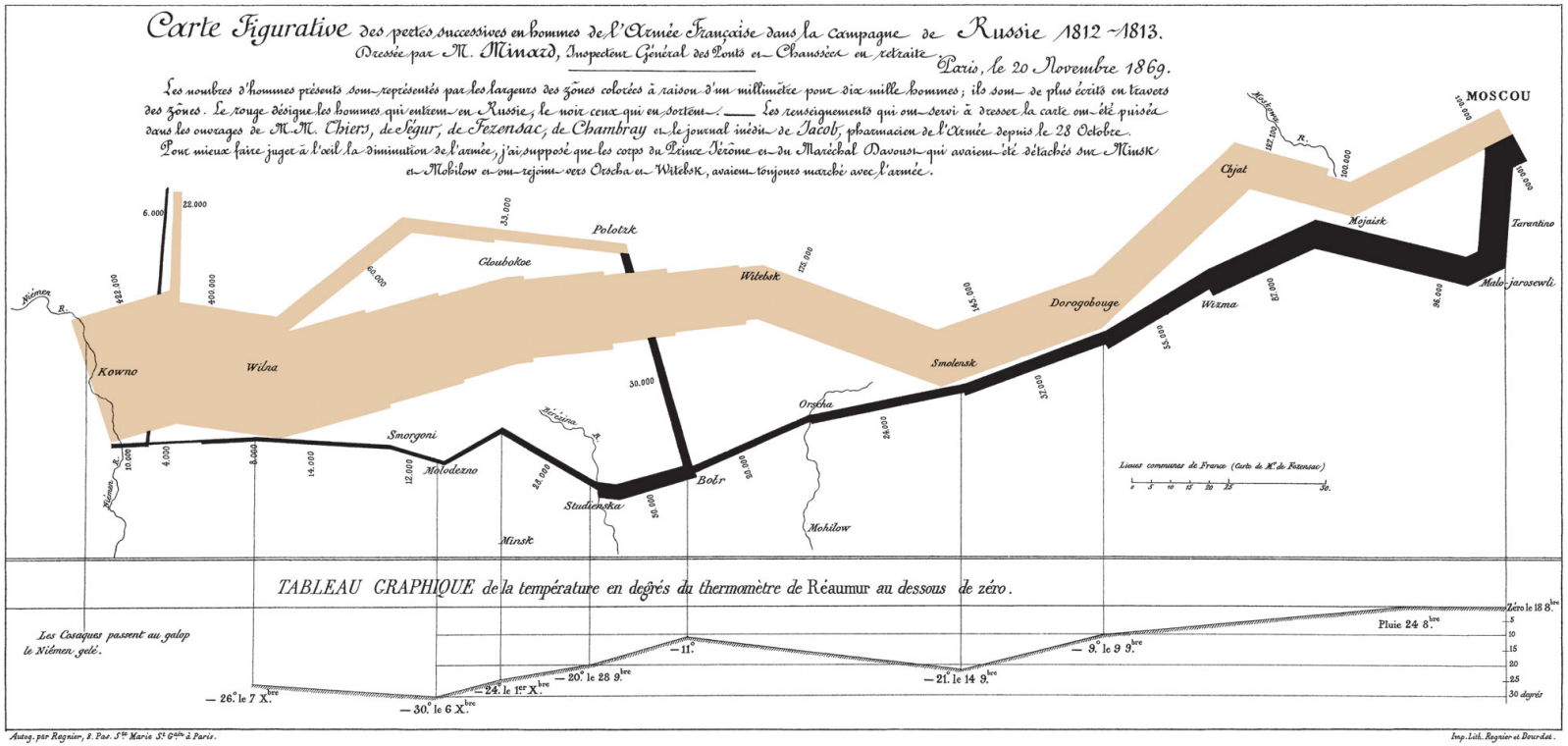
International Trade
Centro de Investigación y Docencia Económicas
January - May, 2023
Data Visualization Project

April 10, 2026

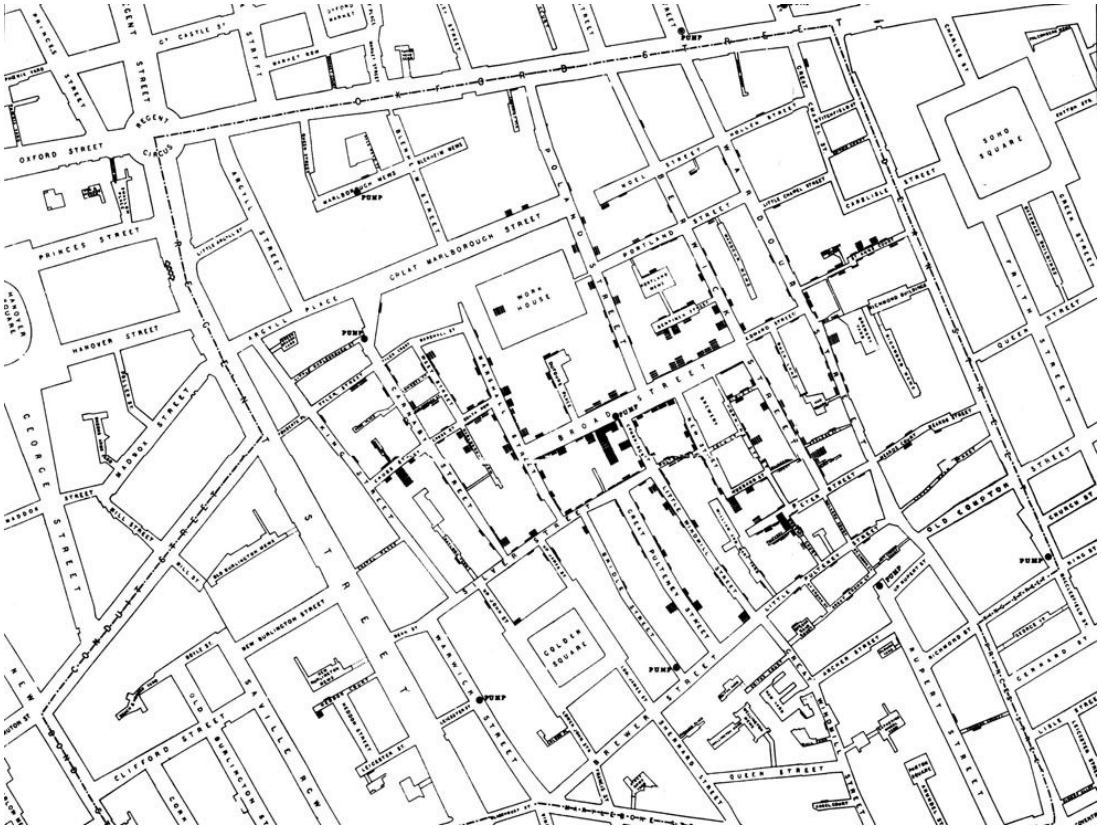
Napoleon's Russian Campaign of 1812 by Minard (1861)



Overview

- Edward Tufte: "the best statistical graphic ever drawn"
- Six types of information
  - Geography: landmarks and battles
  - Time
  - Temperature: for return journey, along the bottom
  - Course and direction: gold and black
  - Troops remaining: one millimeter for 10,000 men
- Source: datavizblog

Cholera Outbreak



Overview

- Cholera in London's Soho :
  - Believed to be spread by unpleasant vapor
  - Germs were not yet understood
  - Sudden and serious outbreak was a mystery.
- John Snow (the original) mapped cases and identified a cluster around the pump in Broad (now Broadwick) street
  - Outliers were asked to link up to the area
  - Close workhouses had own water supply
- Source: The Guardian

Contemporary Masterpiece

Objective

Visualize important data concerning international trade in the widest sense.

Visualization

Visualize important data concerning international trade in the widest sense.

- Important toolkit in debates,
  - big data
  - social media.
- What is data visualization?

  A [data] visualization is any visual display intended to reveal evidence, making the invisible visible.
  Alberto Cairo (2015)

- First part: a definition
  - I will not accept an audio file.
- Second part: an aim

Your goal is to point something out!

Relevancy

Visualize ***important*** data concerning international trade in the widest sense.

- Audience
  - Large
  - Important
- issue
  - Costly
  - Unnecessary
- transmission
  - Understandable display
  - Comprehensible message

International

Visualize important data concerning ***international trade*** in the widest sense.

- Anything that affects cross-country trade.
  - Lecture notes
  - Beyond!

Reach

Visualize important data concerning international trade ***in the widest sense***.

- Theoretical relationship
  - Gravity equation
- Immediate living condition
  - Gender differences
- Business concern

Overview

Date	Time	Topic
17th of February	end-of-day	Consult for topic
22nd of February	classroom	Good viz, bad viz, ugly viz
10th of March	end-of-day	Project topic and data submission
22nd of March	classroom	Rough draft
10th of May	classroom	Final submission

Form teams

- Today!

1. I assign teams of 1-3
2. You assign tasks! You will require
   - Communication and scheduling
   - Literature review
   - Data work
   - Transformation techniques
   - Decide on visualization forms and execute

   Not every tasks needs to be done by one member of your team alone!

Good viz, bad viz, ugly viz

- Find 1 good, 1 bad, and 1 ugly example of data visualizations

1. Write 3-4 sentences for each examples
2. Collect examples and comments in a presentation you give in class before (or on) the due date.

Project topic and data submission

- Follow the next steps:

1. Discuss possible topics
2. Verify that data is available
3. Repeat until you found what you were looking for!

Then,

1. Write a title for your project,
2. Outline the general idea in 3-4 sentences, and
3. Upload the document and the data you will use before (or on) the due date.

Rough draft

1. Apply any data transformation and supporting statistics and plot your data.
2. Present in class before (or on) the due date:
    1. Title of project
    2. Display of data
    3. Message
    4. Targeted audience
    5. Why do you think it is important
    6. Why do you chose the form of visualization you display.
3. Receive feedback!
4. Give feedback!

Final submission

- Email PNG file before (or on) the due date
- All projects
  - will be discussed in class
  - graded by your classmates

Half of your final grade for this project will depend on their grade!

Grading criteria

Dimension	Exemplary	Unsatisfactory	Weight
Title	descriptive, informative, intriguing	no title	10%
Description	efficient use of text elements	display is inexplicable	10%
Clarity	data is identifiable immediately	choices lead to ambiguous display	10%
Support	assist visualization	data clutter	20%
Accuracy	data is accurately displayed	display deceives the audience	10%
Appropriateness	display matches message	display does not mirror message	20%
Impact	could headline an Economist article	is a waste of time and attention	20%
Grade	10	6	100%

General advice I

1. Tell a story.
2. "Show the data" $\not\Rightarrow$ "show *all* the data".
   - Declutter, and avoid a "data dump".
3. It is easier to identify objects by color than by shape.
   - Use color intentionally.
   - black-white printouts.
4. Don't graph for the author!
   - Data exploration vs conveying a message
5. Default graph options in software tends to clutter and separate text and graphs
   - Add annotations and avoid legends
   - Researcher explore using default grid lines, tick marks, and colors
6. Elements such as text, lines, shapes, etc, that have the same nature should look alike.
7. Strive for balance between graphic and explanatory text.

General advice II

8. Prevent your audience from asking "so what?" by showing
   - contrasts,
   - comparisons,
   - differences.
9. Avoid the third dimension, it is usually not helpful.

Specifics I

1. Data sources should be included in a very brief way.
2. You can use supporting statistics such as filters, trends, etc., to make your point.
3. I provide some examples with Matlab code.
   - There is no software requirement.
4. Some tools and data sources are mentioned in the document "Data visualization assignment" on the webpage of this course.
5. The aforementioned points are suggestions based on my own taste and experience about prevalent usage. They are not commandments.

Count the 3s

1269548523612356987458245

0124036985702069568312781

2439862012478136982173256

Count again

126954852**3**6123569874582450124**3**6985702069568**3**12781243986201247813698217**3**256

Pre-attentive visual processing

- Attentive processing
- Pre-attentive processing is done in parallel and faster

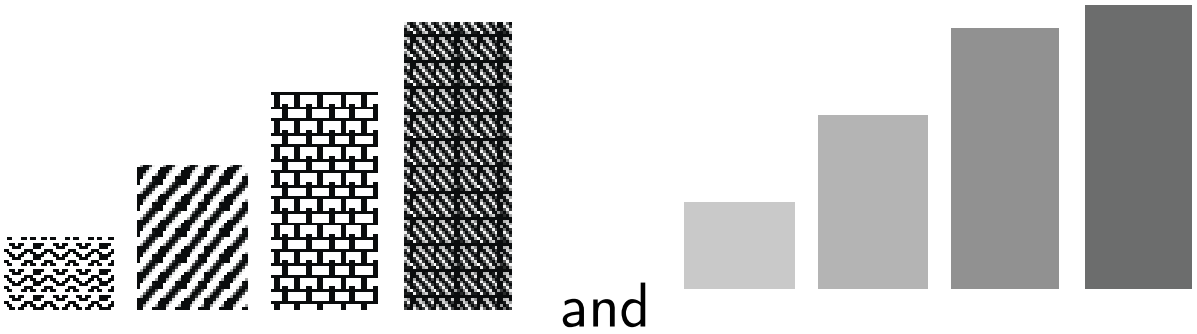
Pre-attentive visual processing

- Attentive processing
  - the conscious part of perception that allows us to perceive things serially
- Pre-attentive processing is done in parallel and faster
  - instances of 3s are easier to find because they are encoded using a different pre-atttentive attribute
  - intensity of boldface type.

Picture Superiority Effect

- Our ability to retain information from pictures exceeds the ability to do the same through words

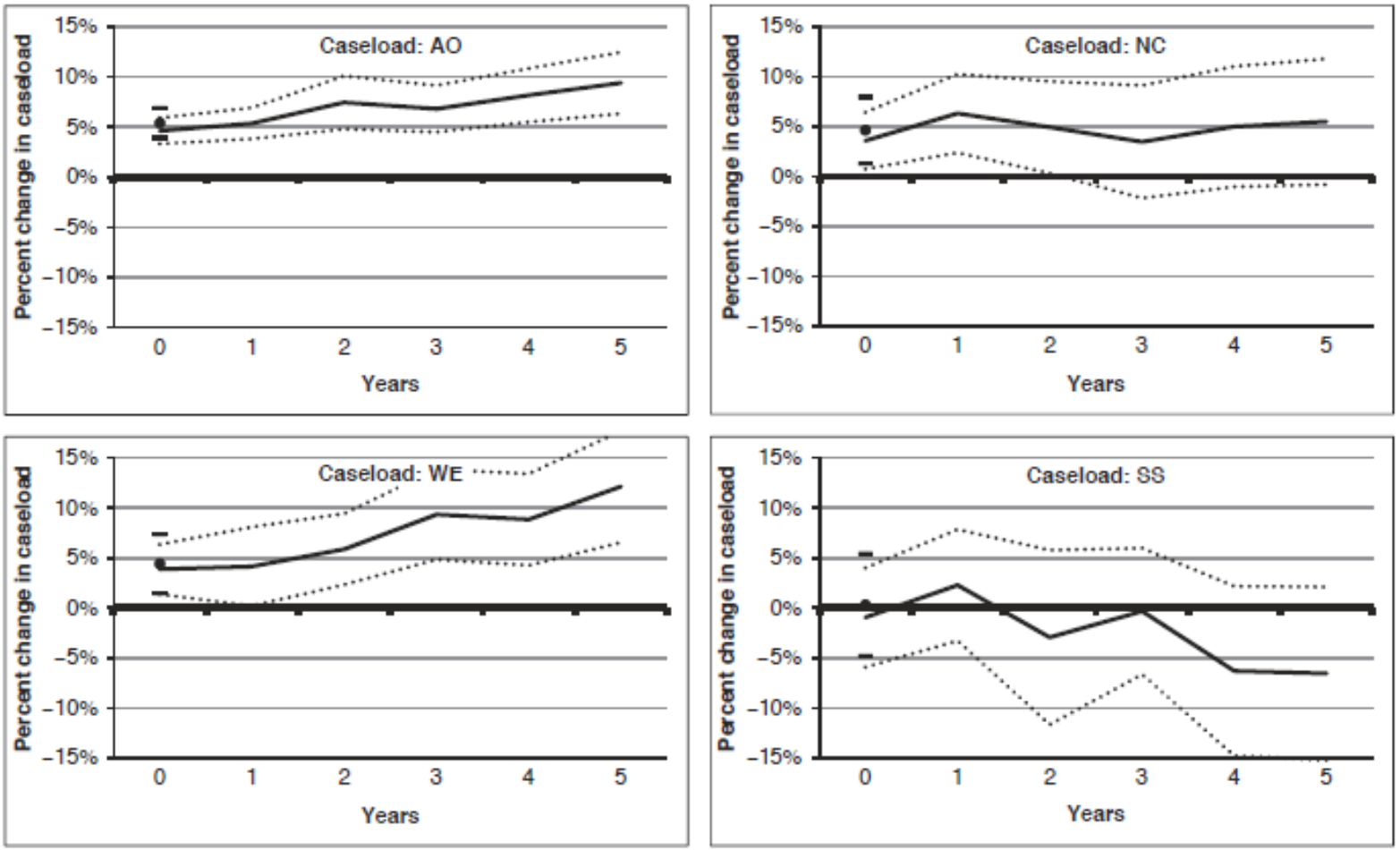
Declutter



and

display the same information.

Original



Source: Klerman and Danielson (2011).

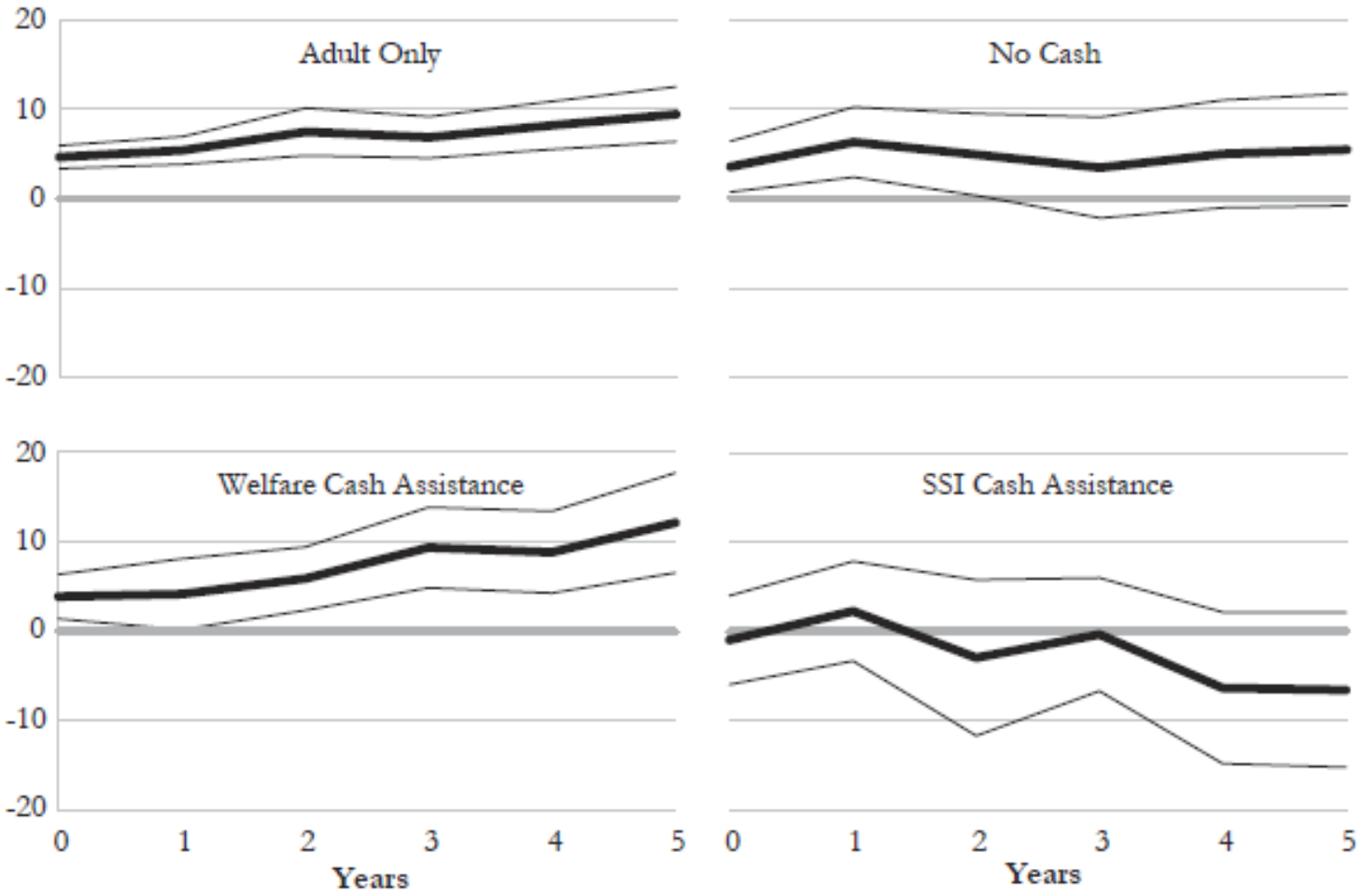
Message

- Klerman and Danielson (2011)
- Message: regression results of the correlation between the long-run unemployment rate in the United States and Supplemental Nutrition Assistance Program caseloads for four groups
- Instead of single line: four smaller charts combined
  - Not a bad idea

Critique

1. The darkest and thickest line is the 0 percent grid line.
   - not the data!
2. Some data values exceed the upper bound of the chart
3. Clutter:
   - The y-axis labels and percentage signs are redundant
     - 28 percentage signs in total!
   - Tick marks on the y-axes are unnecessary.
4. What do AO, NC, WE, and SS mean?
   - Terms are explained on the third and fourth pages, 15 pages before this figure

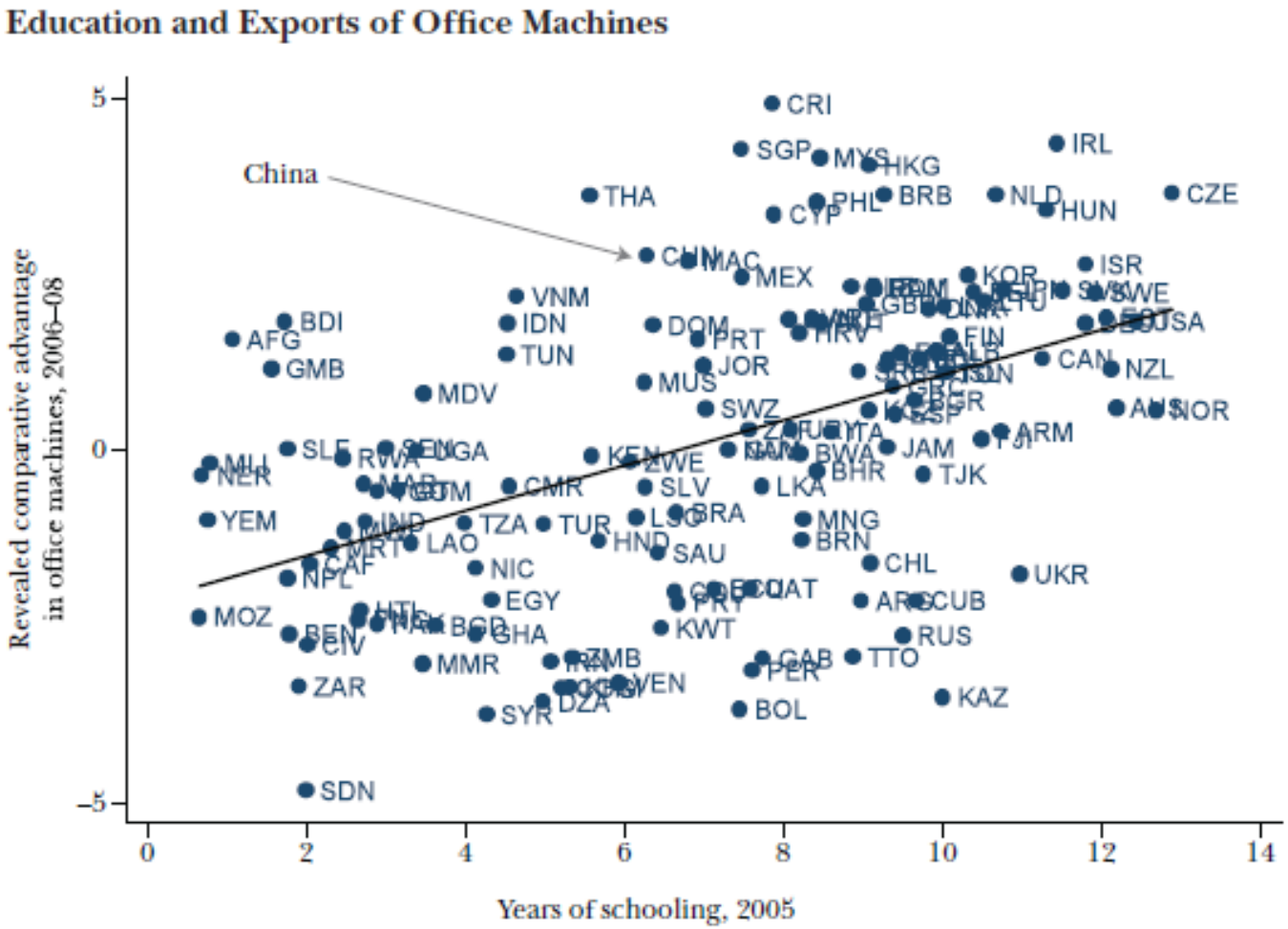
Implied Impulse Response Functions for Different Caseloads
(Percent change)



Improvement

- The darkest line shows the data
- The grid lines are lighter
  - leaving the 0 percent grid line darker as a reference.
- Two sets of labels are gone
  - The charts are aligned vertically and horizontally
- Percent signs are gone
  - Units are below the title
- y-label moved to the title
  - Rotated text on the vertical axis requires readers to turn the page sideways or tilt their heads
- The title is now above the graph
  - Readers tend to start reading from the top left, move down along the left margin, and then move to the right.
- "Caseload" moves into the title
  - used four times in the original
- Abbreviations are spelled out

Original



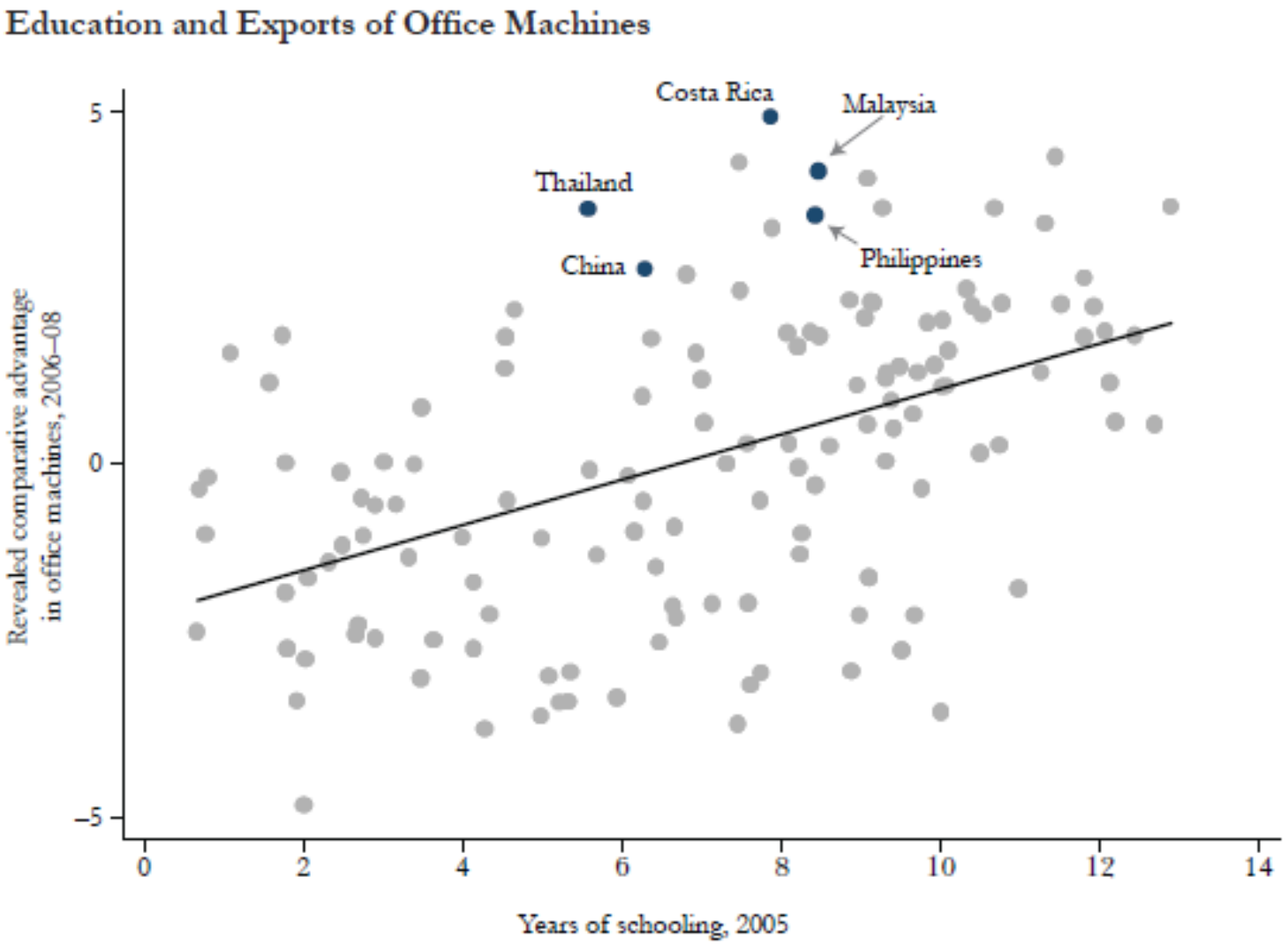
Source: Hanson (2012).

Message

- Hanson (2012, JEP)
- "[The figure] plots countries' revealed comparative advantage in office machines . . . averaged over 2006 to 2008, against the average years of schooling of the adult population in 2005 . . . China is above the regression line, indicating that its specialization in the sector is greater than one would expect given its level of education, but it is hardly an extreme outlier. Other middle-income countries—including Costa Rica, the Philippines, Malaysia, and Thailand—have larger positive residuals."

Critique

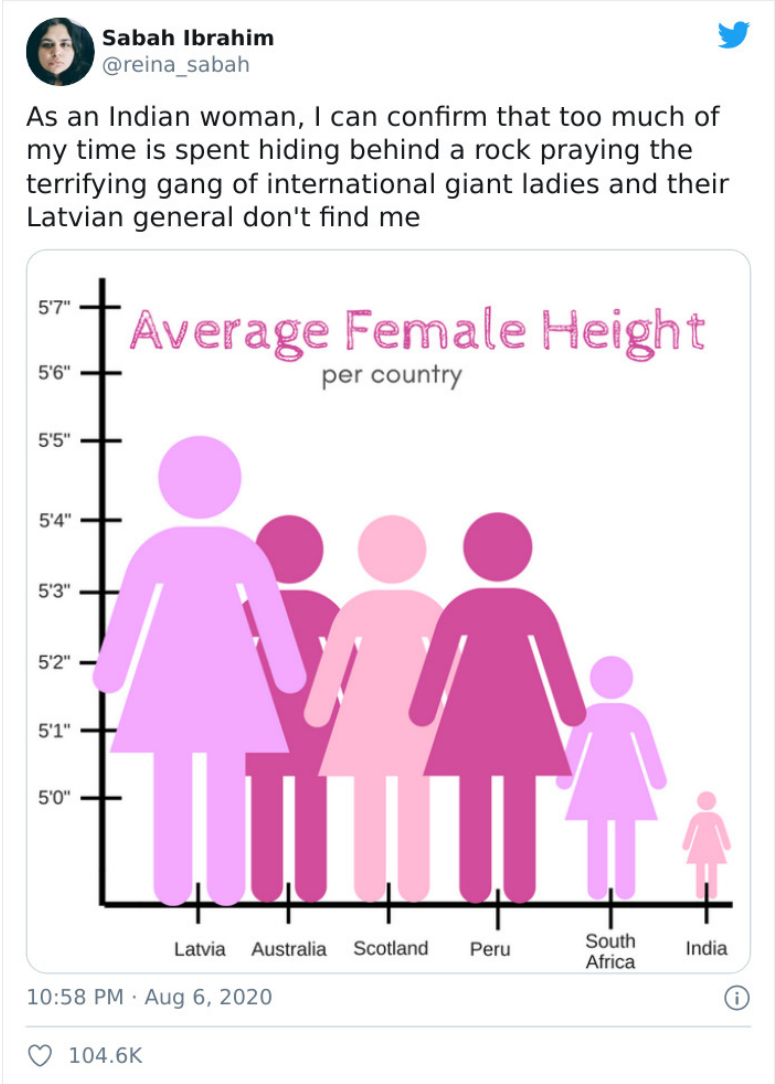
- Do you know the 3 letter code for China, Costa Rica, the Philippines, Malaysia, and Thailand?


Education and Exports of Office Machines
Costa Rica
Malaysia
Thailand
China
Philippines
5
0
−5
Revealed comparative advantage in office machines, 2006–08
0
2
4
6
8
10
12
14
Years of schooling, 2005

Improvement

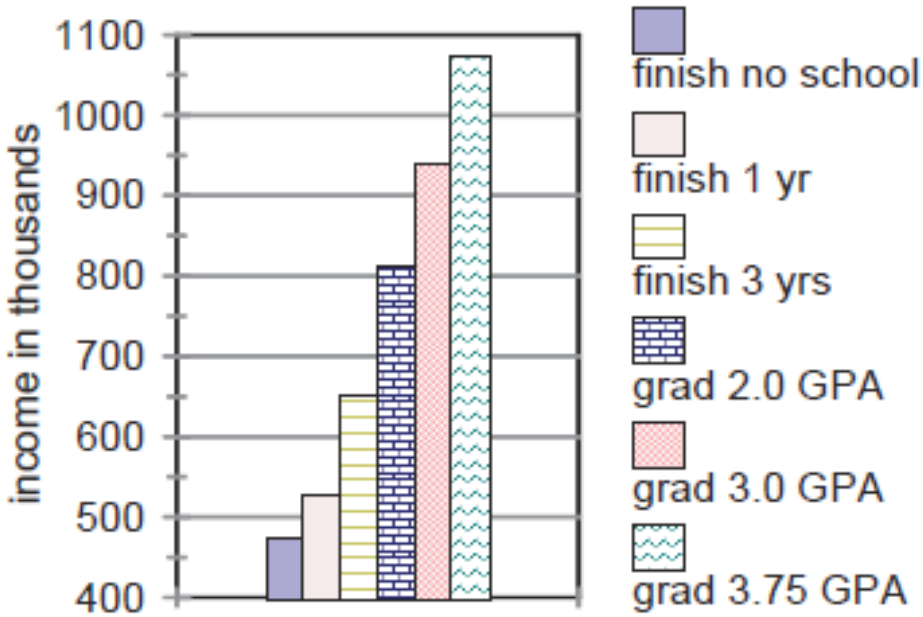
- Eliminate all but 5 relevant countries and spell out those.
  - Conflict: some reader might want to know about individual countries or identify outliers.
    - Classic trade-off
    - supplemental table or data provision

Intermezzo


Sabah Ibrahim
@reina_sabah
As an Indian woman, I can confirm that too much of my time is spent hiding behind a rock praying the terrifying gang of international giant ladies and their Latvian general don't find me
Average Female Height
per country
5'7"
5'6"
5'5"
5'4"
5'3"
5'2"
5'1"
5'0"
Latvia
Australia
Scotland
Peru
South Africa
India
10:58 PM · Aug 6, 2020
104.6K

Original

Figure 2 Discounted Expected Lifetime Earnings, VN(t')



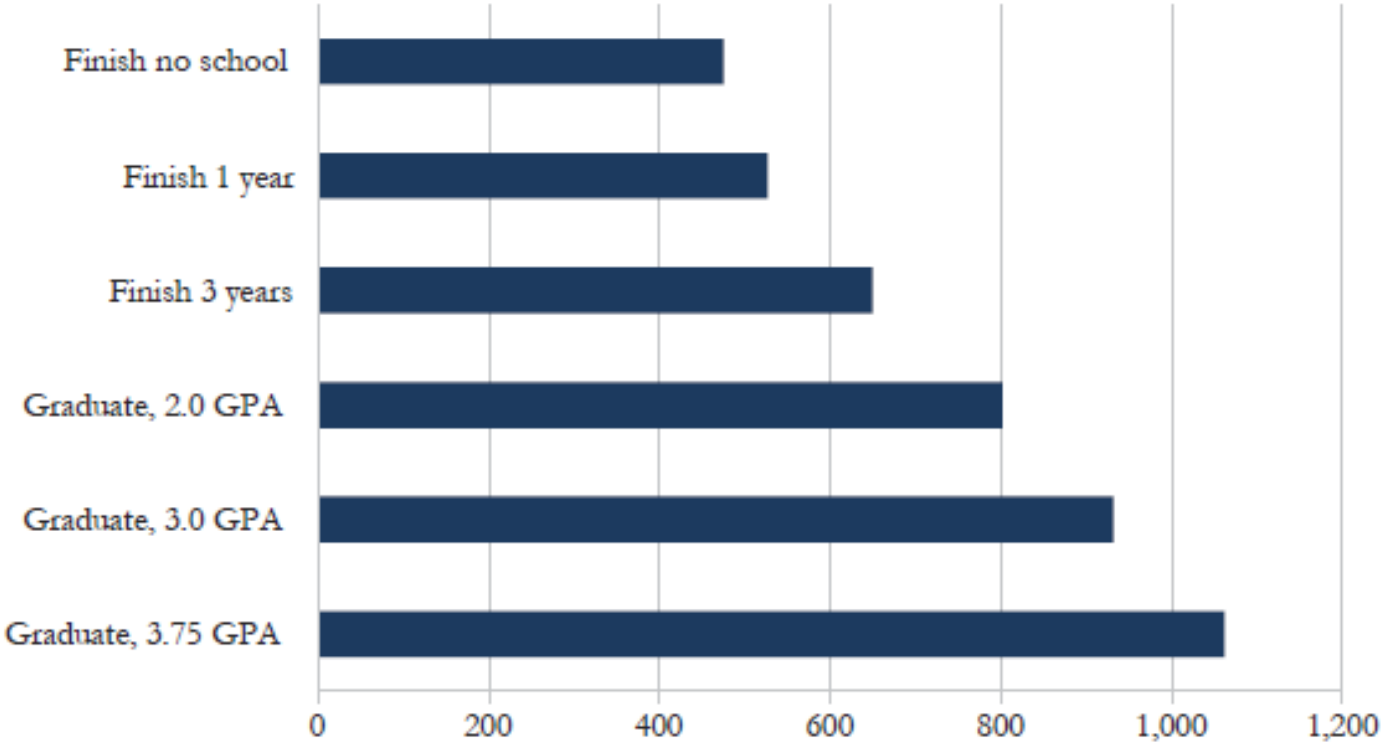
Source: Stinebrickner and Stinebrickner (2013).

Are discounted expected lifetime earnings of someone with 1 year of schooling half of those finishing with a 3.75 GPA?

Critique

- My eyes say no!
- Different colors/patterns for each bar - why?
  - visual matching using a legend

Discounted Expected Lifetime Earnings, VN(t')
(Income in thousands)



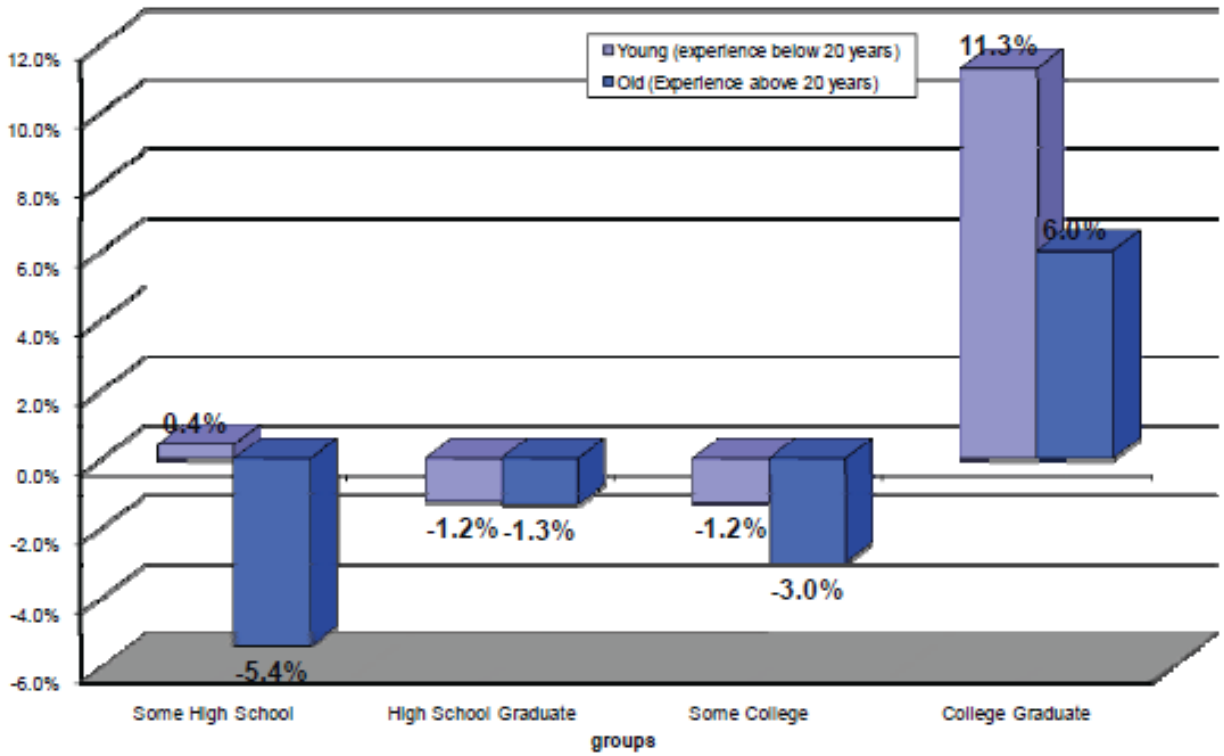
Source: Author's calculations using numbers inferred from text in Stinebrickner and Stinebrickner (2013).

Improvement

- The axis starts at zero
- The figure is rotated horizontally
  - to make room for the labels
- Finish 1 year vs grad 3.75 GPA appear comparable

Original

Change in real weekly wages of US-born workers by group, 1990-2006

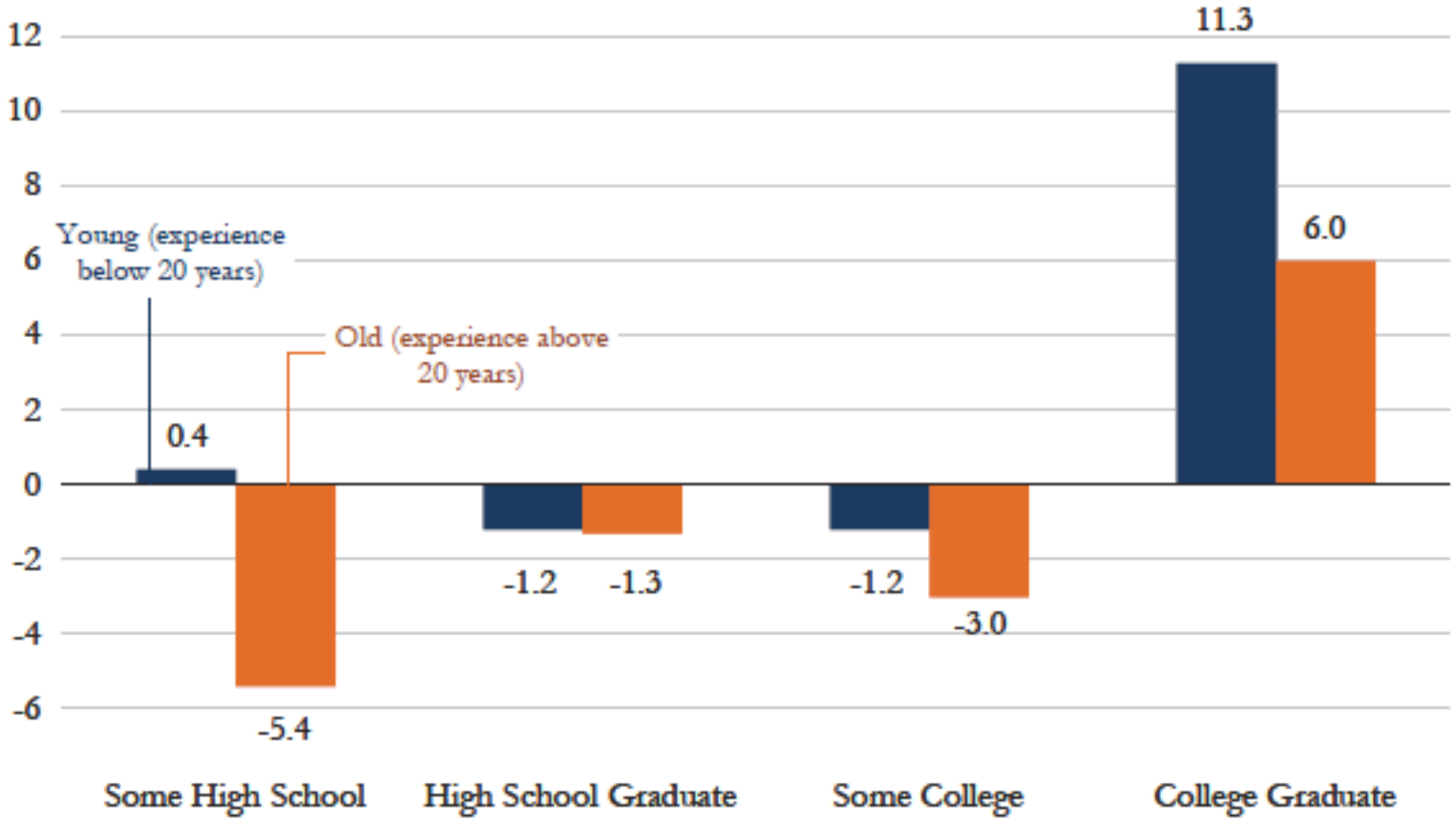


Source: Ottaviano and Peri (2008).

Critique

- Third dimension
  - does not plot data values
  - adds clutter to the chart
  - distorts the information.
    - far-right-hand bar, labeled 6 percent, does not touch gridline

Change in real weekly wages of US-born workers by group, 1990–2006
(Percent)

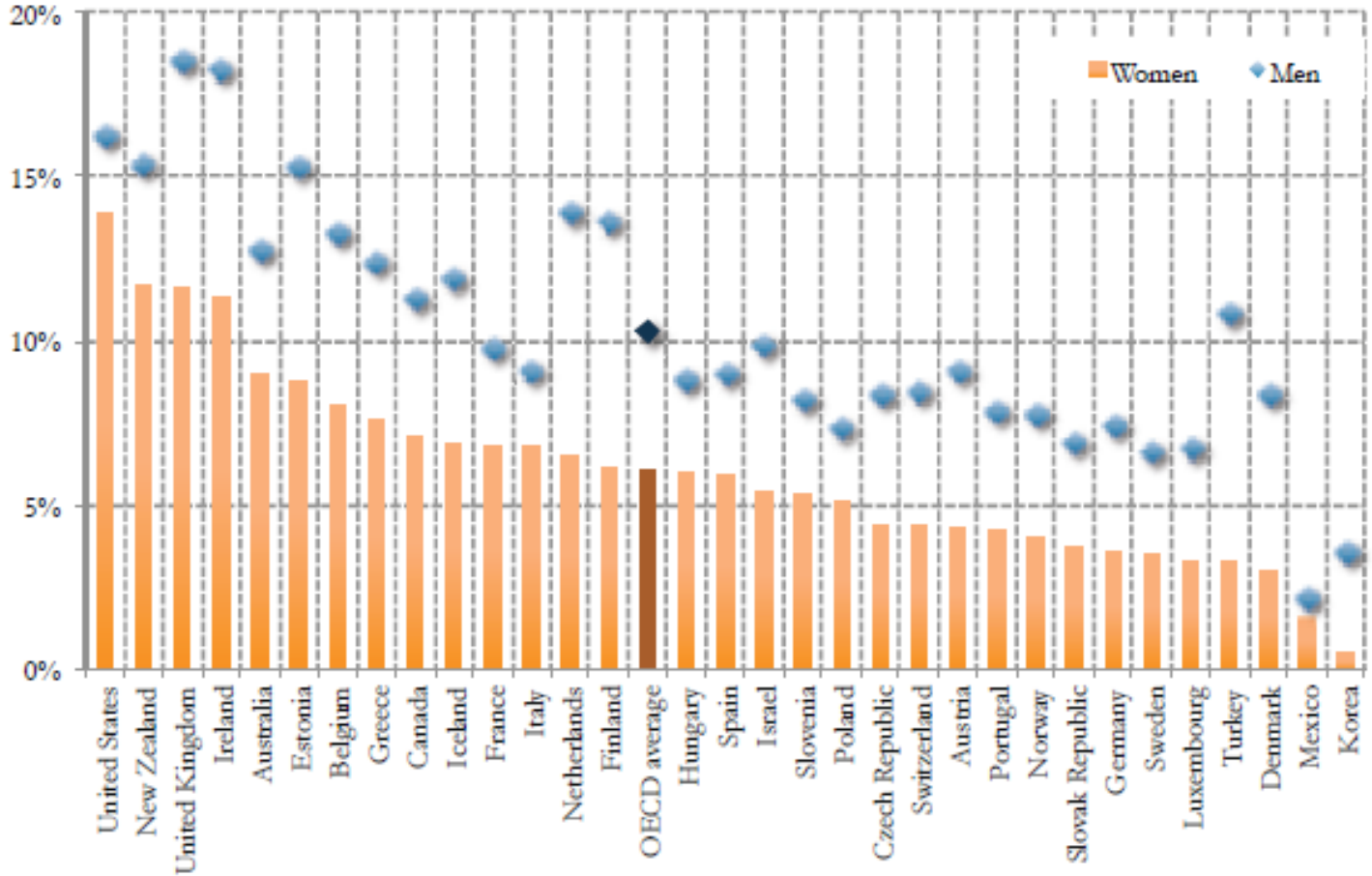


Improvement

- Cancel the 3D treatment
- Integrate the disconnected legend with the graph.
- Common baseline permits an effective comparison
  - was portrayed by a hovering, barely perceptible thin gray line

Original

Percentage of Employed Who Are Senior Managers, by Sex, 2008



Source: Author, based on OECD (no date) and Rampell (2013).

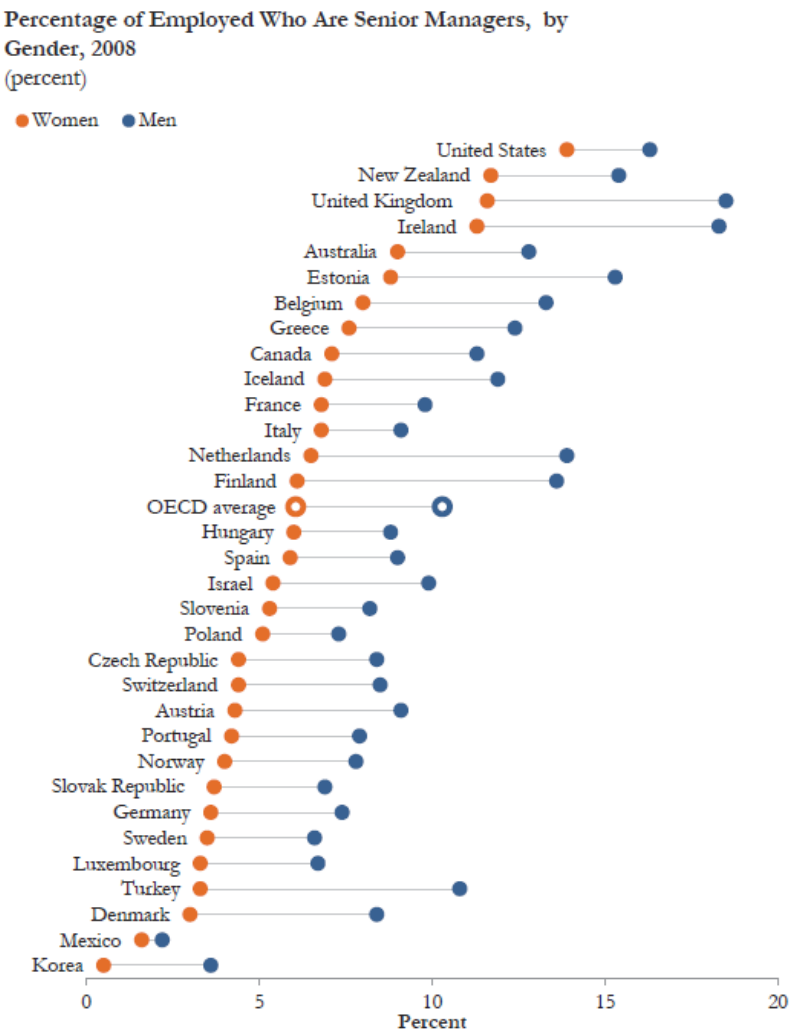
Source

- Organisation for Economic Co-operation and Development
- later reproduced in a New York Times Economix blog post

Critique

- Same kinds of data are plotted using different types of encoding
  - diamonds and bars
- The men are too far away from the women without visual connection
- The columns for women take up a large proportion, overemphasizing the data for women.
  - Focus on women? Then change the title: "Women's Employment as Senior Managers Averaged 6 Percent in 2008"
- Too many grid lines
- The percent signs on the y-axis labels are redundant.
- X-axis labels are potentially difficult to scan because they are vertical.

Revision

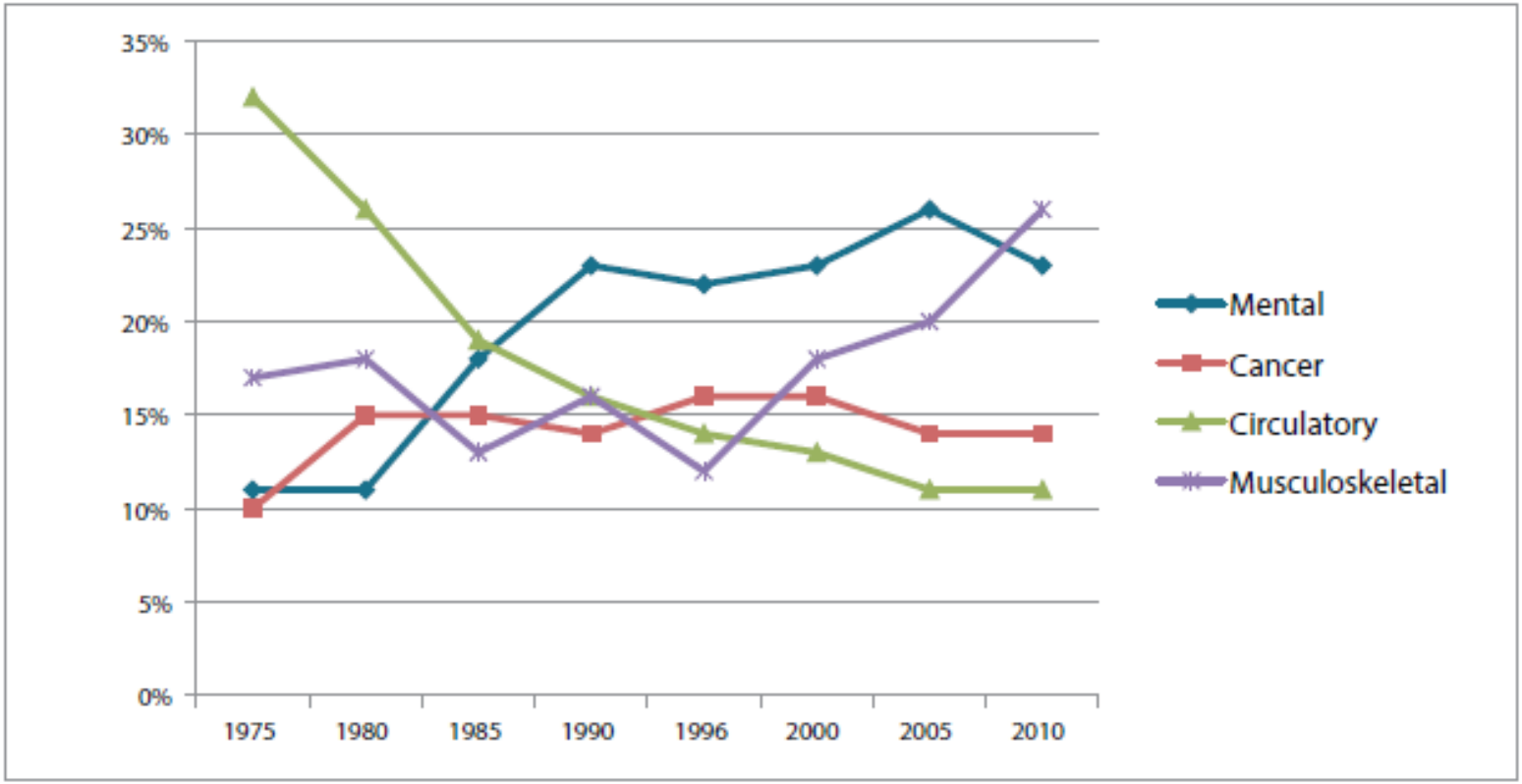

Percentage of Employed Who Are Senior Managers, by Gender, 2008
(percent)
Women
Men
United States
New Zealand
United Kingdom
Ireland
Australia
Estonia
Belgium
Greece
Canada
Iceland
France
Italy
Netherlands
Finland
OECD average
Hungary
Spain
Israel
Slovenia
Poland
Czech Republic
Switzerland
Austria
Portugal
Norway
Slovak Republic
Germany
Sweden
Luxembourg
Turkey
Denmark
Mexico
Korea
0
5
10
15
20
Percent

Improvement

- Visualization a bit "exotic"
- Data is encoded similarly for men and women
  - dots
- Thin gray lines connecting the two series enhances comparison
  - men vs women
  - different countries
- The title, units, and legend are summarily placed at the top-left so the reader can "enter" the chart there.
- The country labels are rotated horizontally and incorporated next to the observation
- The average value for the OECD as a whole is an unfilled circle
  - helps printing gray scale
- Shortcoming: no vertical lines!

Original

27. Initial DI Worker Awards by Major Cause of Disability—Calendar Years 1975-2010

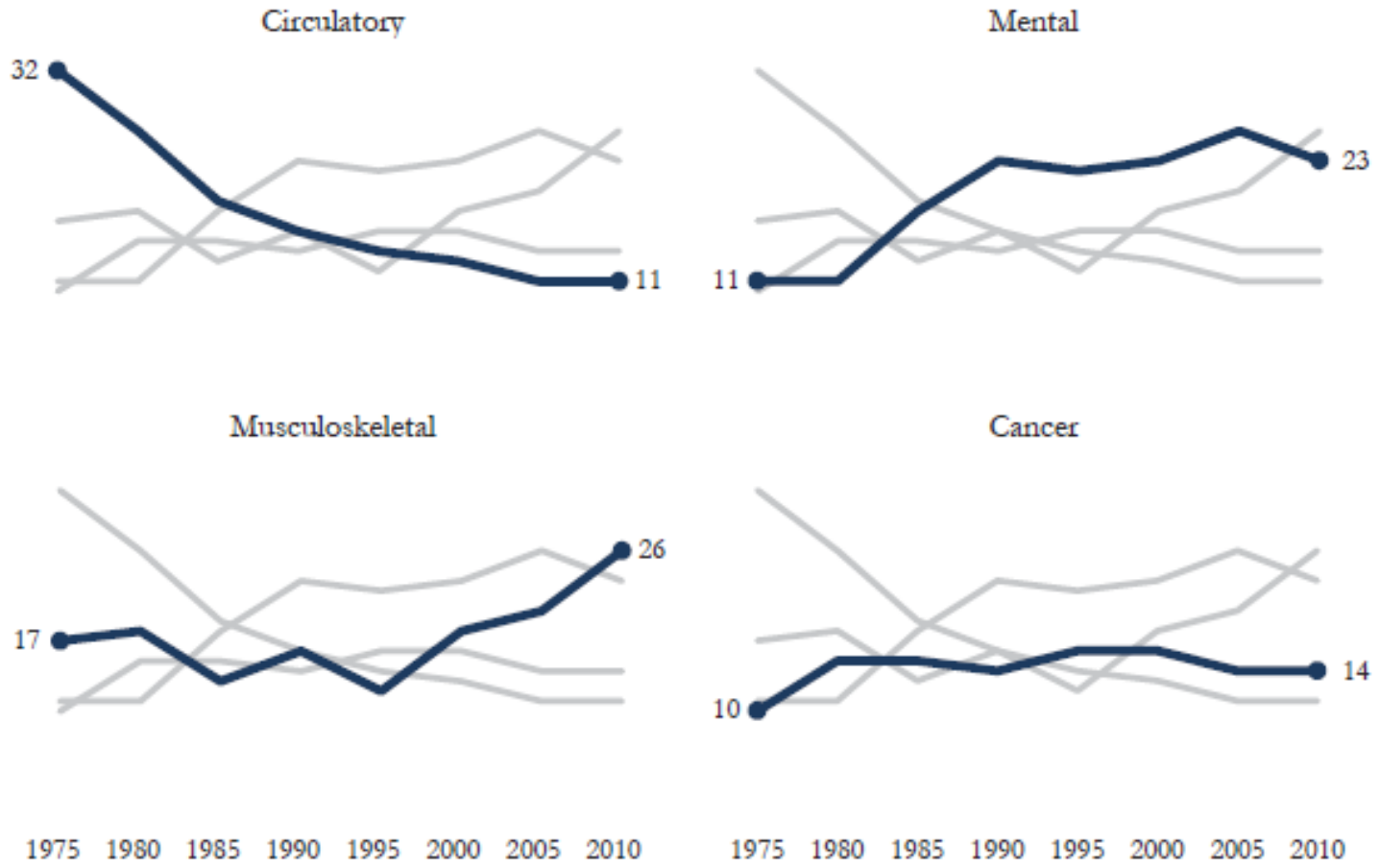


Source: Social Security Advisory Board (2012).

Critique

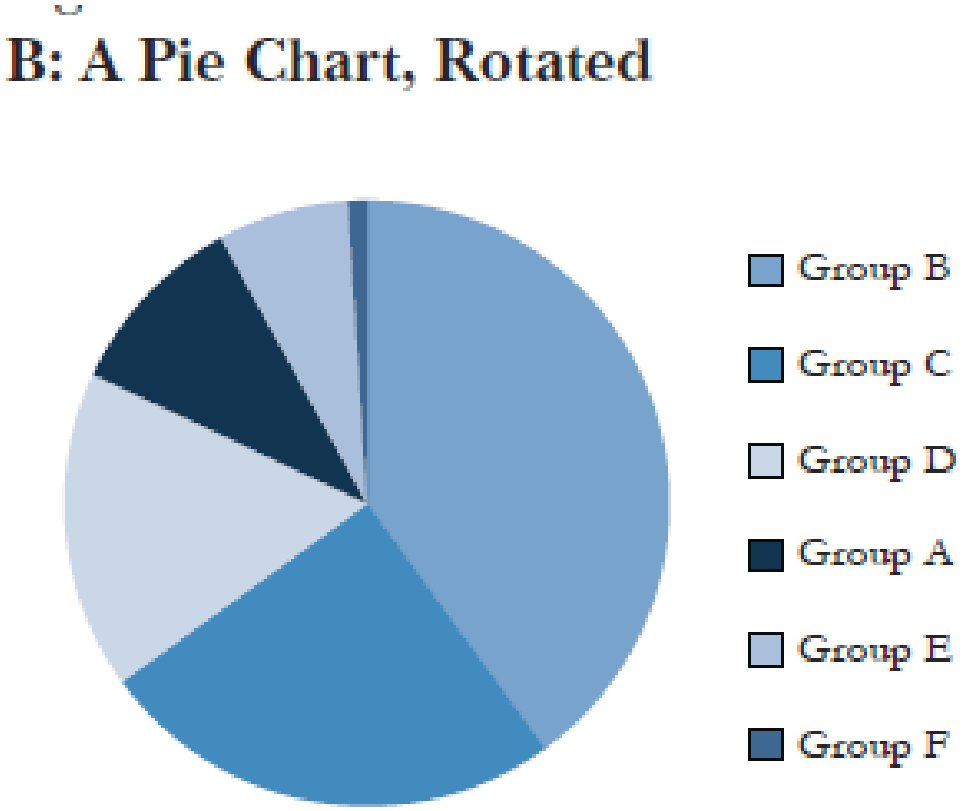
- the legend
  - is far from the data
  - the order of the legend does not match the order of the lines.
- Spaghetti charts, in general, overwhelm if it contains more than, say, 4 lines.

Initial DI Worker Awards by Major Cause of Disability—Calendar Years 1975–2010
(Percent)



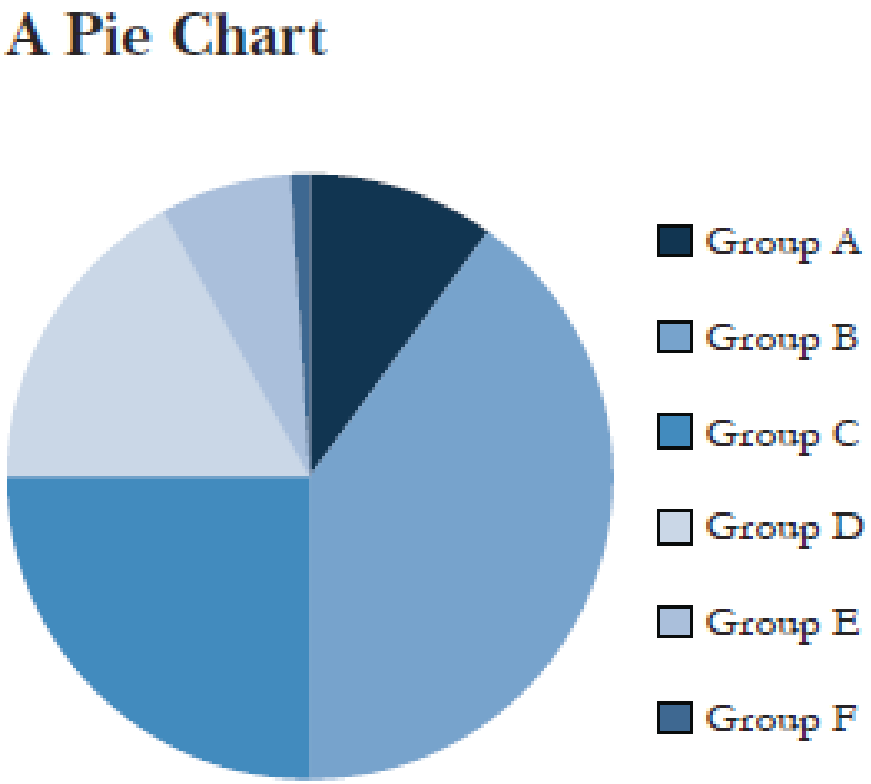
Improvement

- Create smaller charts in series
  - Highlights the trend in each line
  - Without loosing comparison
  - While reducing clutter
- Recycle y-axis labels



What is the fraction of group C?

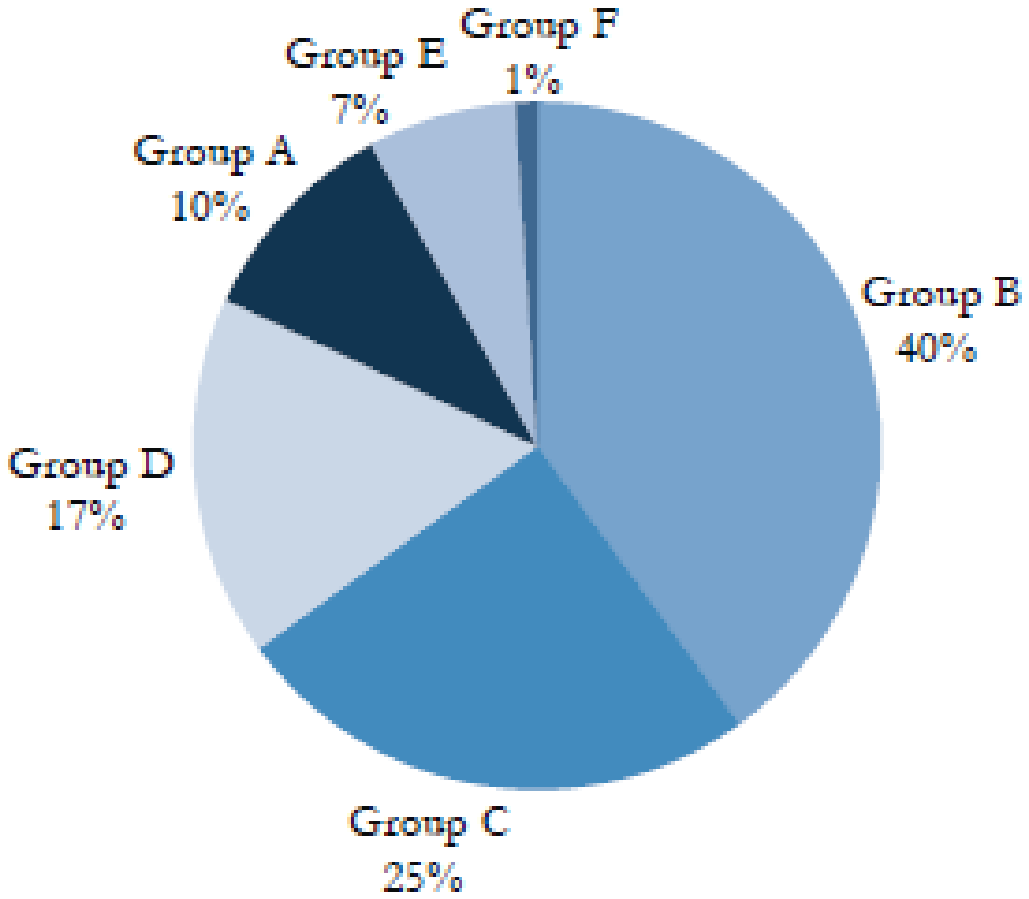
Alphabetical order



Obviously 25%! So, order matters!

What about the rest?

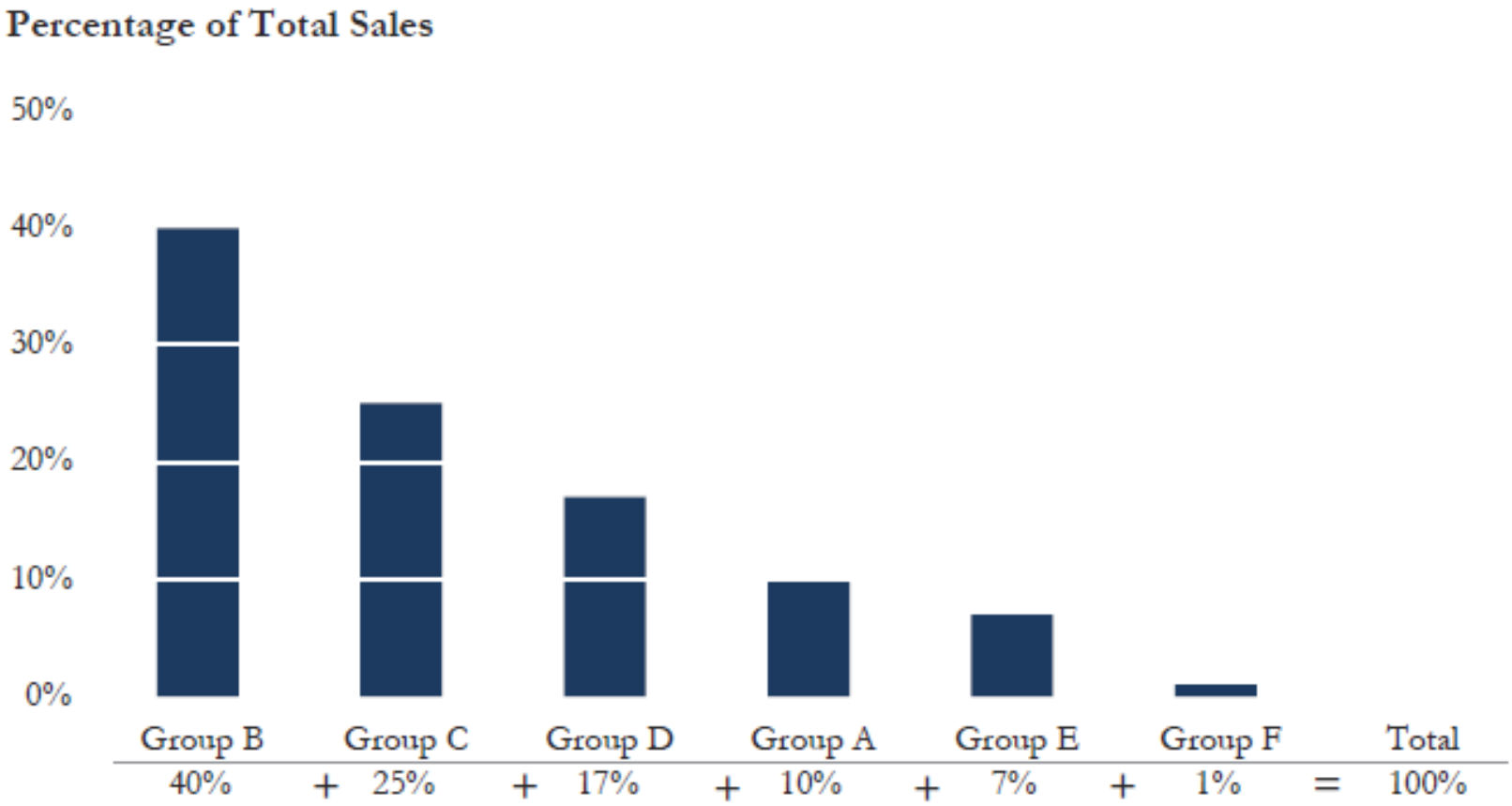
Size, labeled



Statistical pie is not apple pie is not

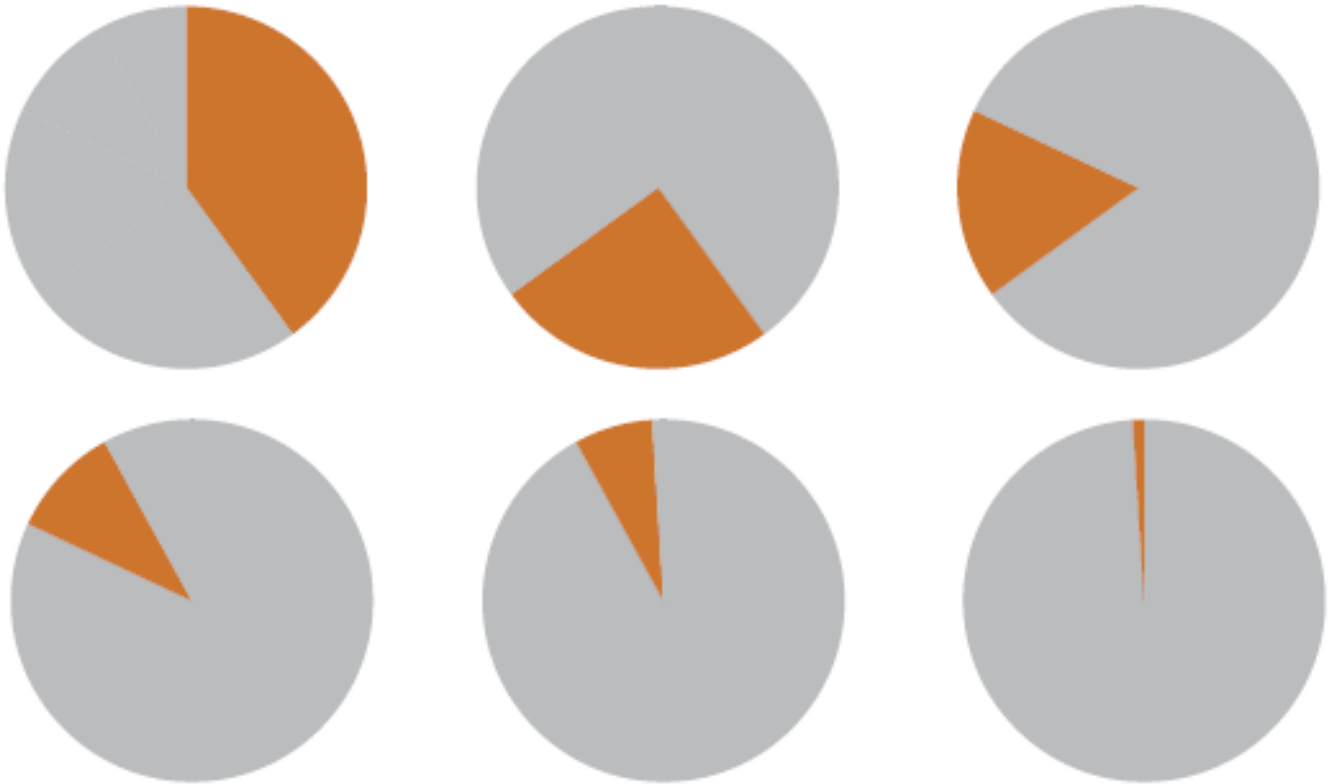
- Comparison of areas is actually difficult
  - worse for Donut chart where center (with angles) is missing

Alternative



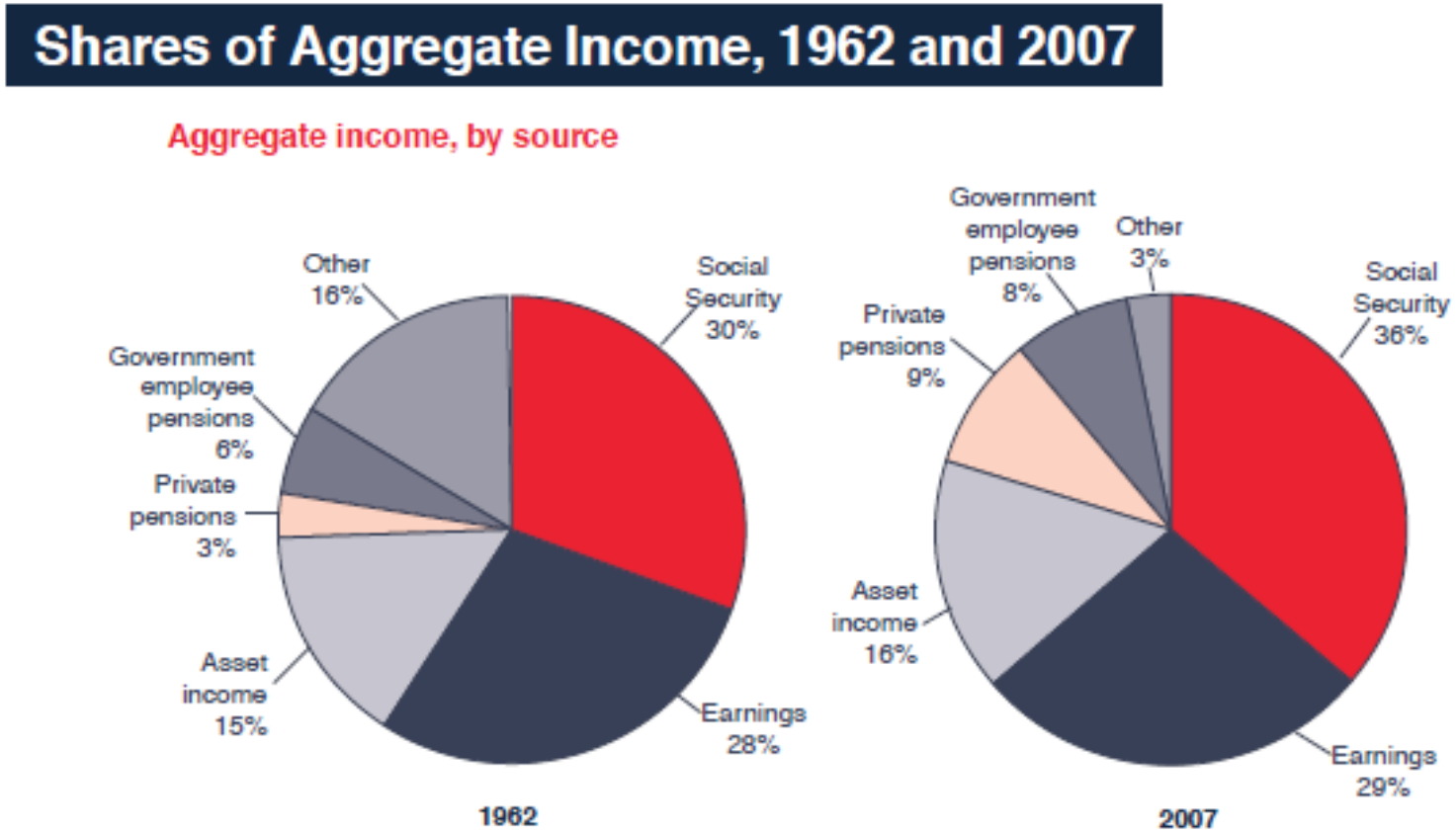
Size comparison immediately accessible.

Alternative 2



- recover "part-of-a-whole" comparison while size-comparison is preserved

An original pie chart



Source: Social Security Administration (2009).

Intention

- Comparison not just
  - part-of-a-whole
  - between different parts

  but also over time!

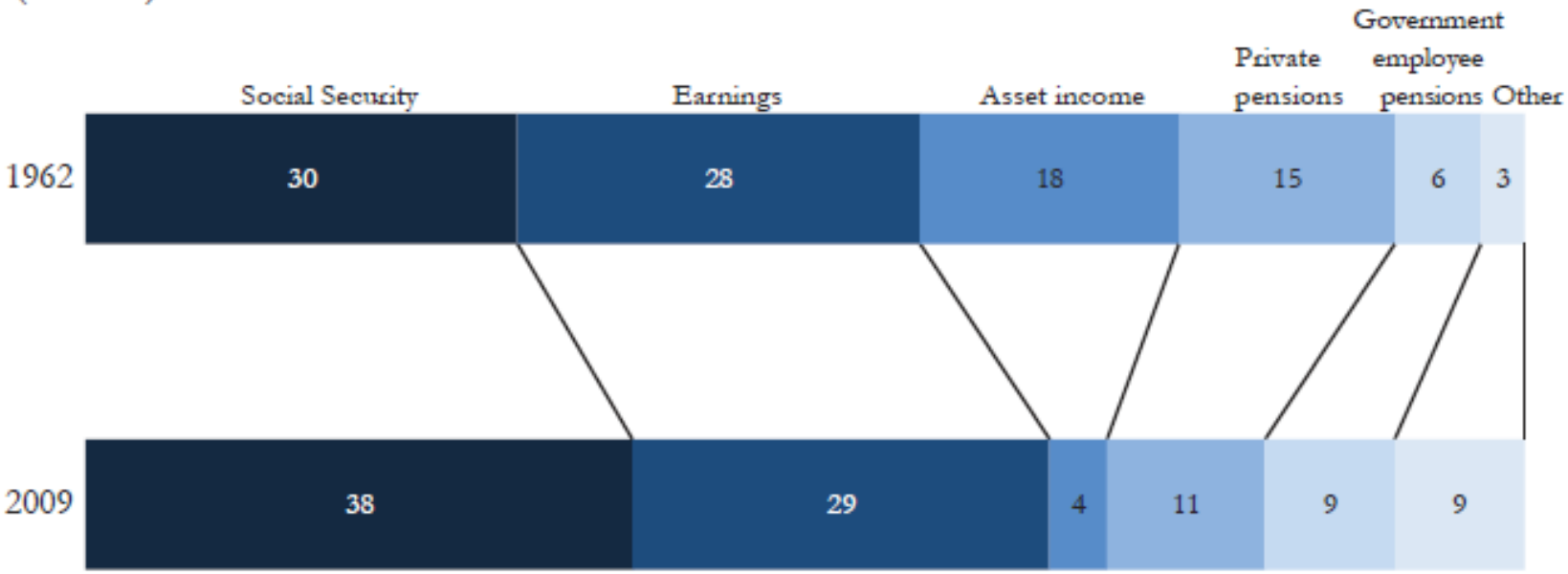
Shares of Aggregate Income, 1962 and 2009
(Percent)

Improvement

- This is called *paired column chart*
  - fosters within-category/over-time comparisons.
- Y-axis is omitted
- Few observations allow for vertical orientation

Another revision

Shares of Aggregate Income, 1962 and 2009
(Percent)



Improvement

- This is called *stacked bar chart*
- Shows
  - the distribution of the various groups
  - the groups sum to 100 percent
  - changes over time

Yet another revision

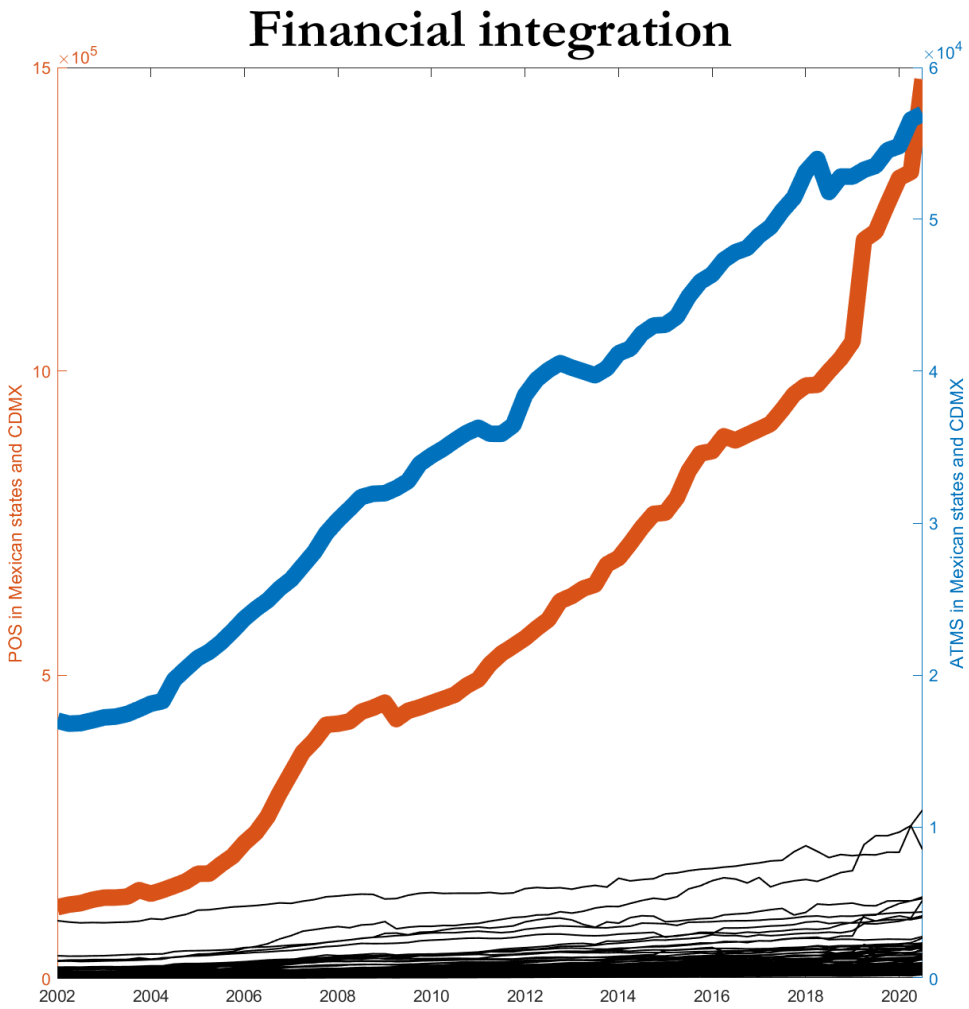
Improvement

- This is called *slope chart*
- Shows
  - the size-order
  - changes over time

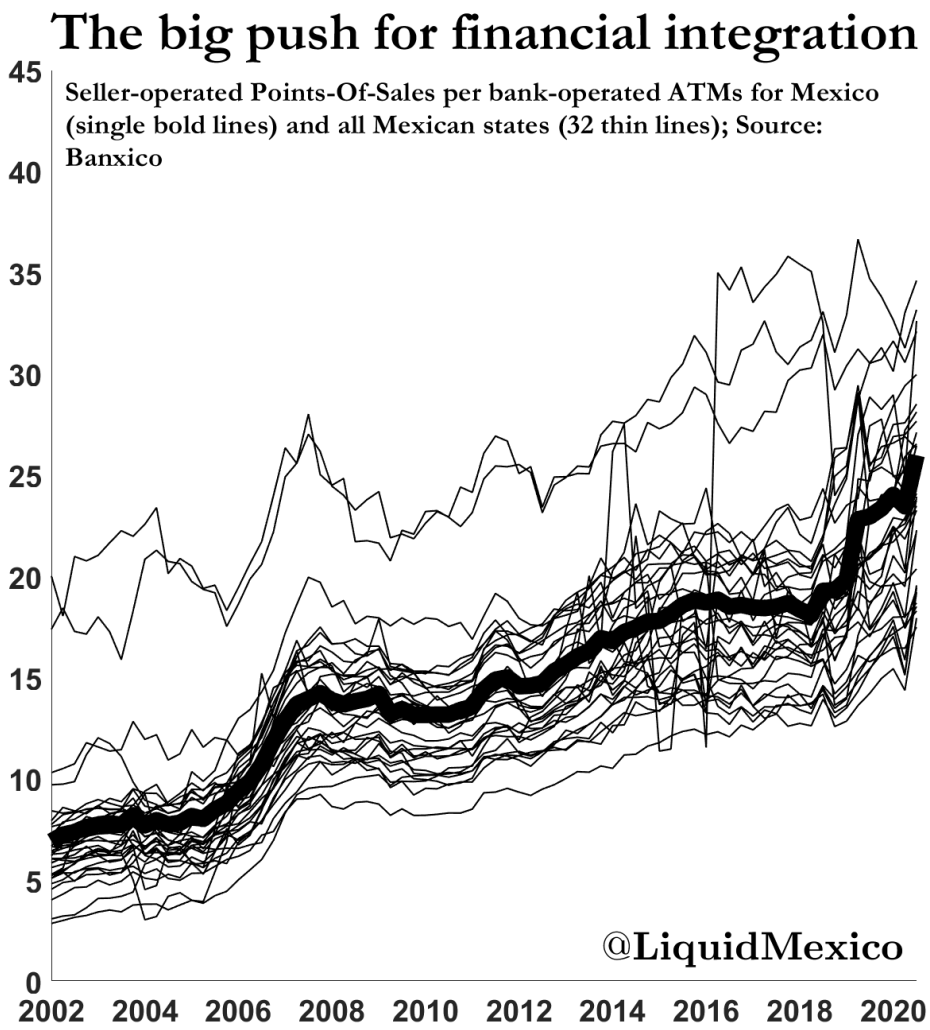
My own motivation

- I would like to engage more over social media
- But note: my examples are very low tech in terms of support!

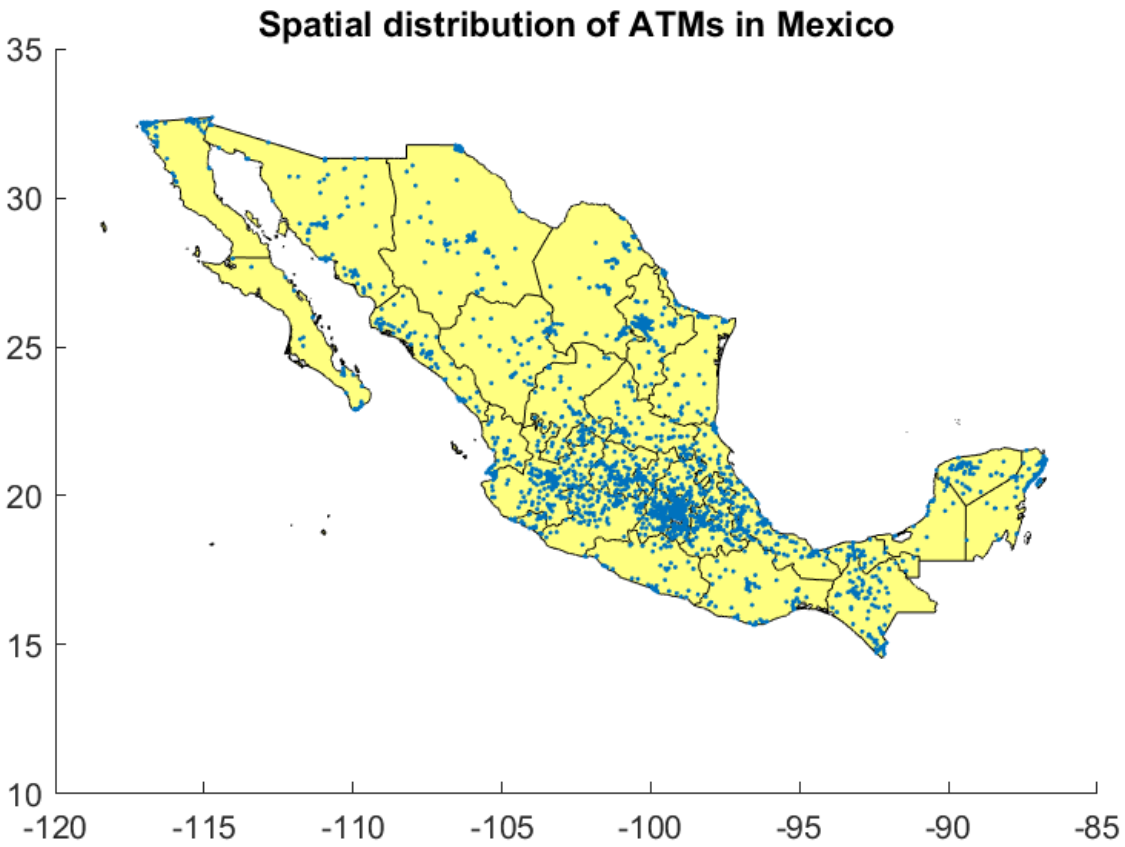
Lets try one:


Financial integration
×10^5
15
10
5
0
POS in Mexican states and CDMX
×10^4
6
5
4
3
2
1
0
ATMS in Mexican states and CDMX
2002
2004
2006
2008
2010
2012
2014
2016
2018
2020

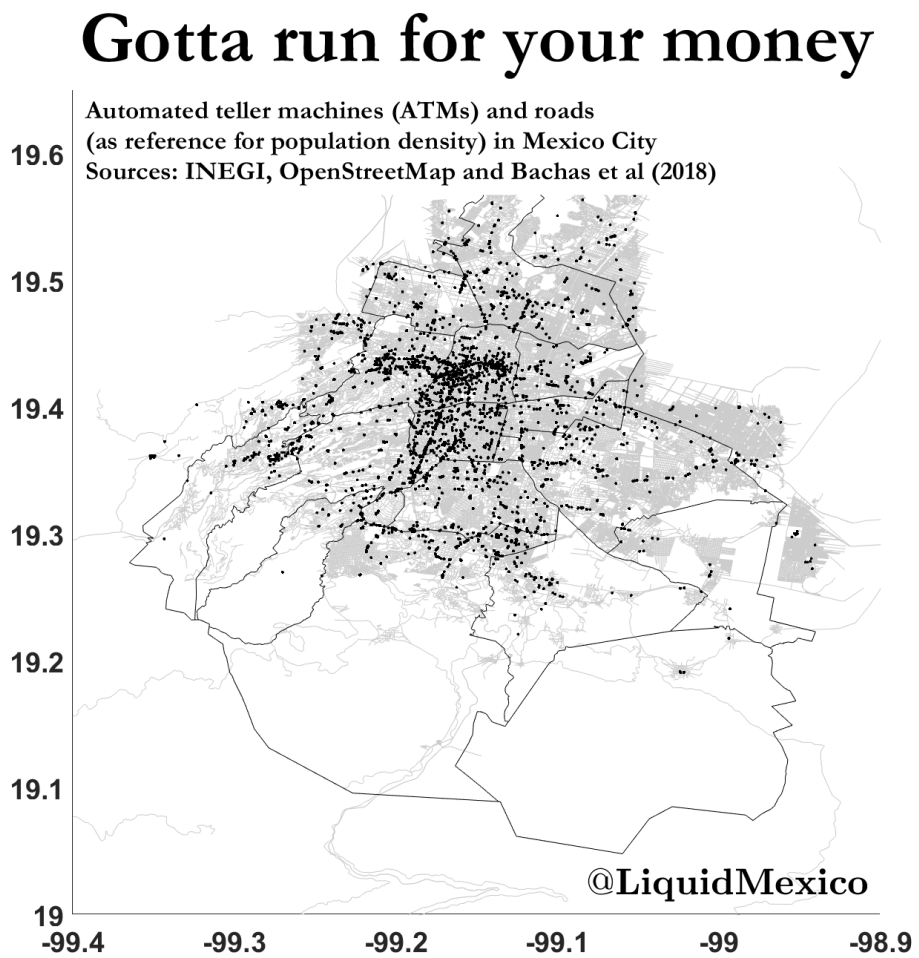
An improvement



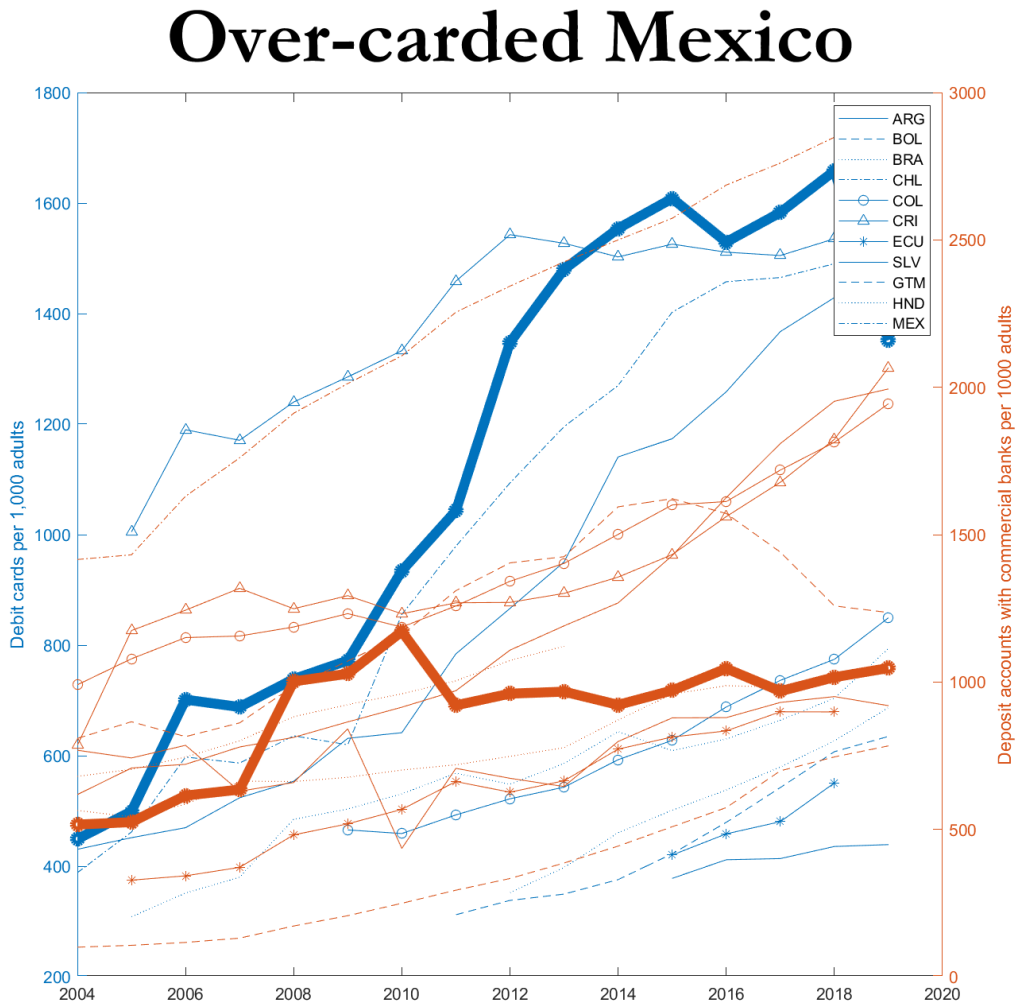
Another attempt



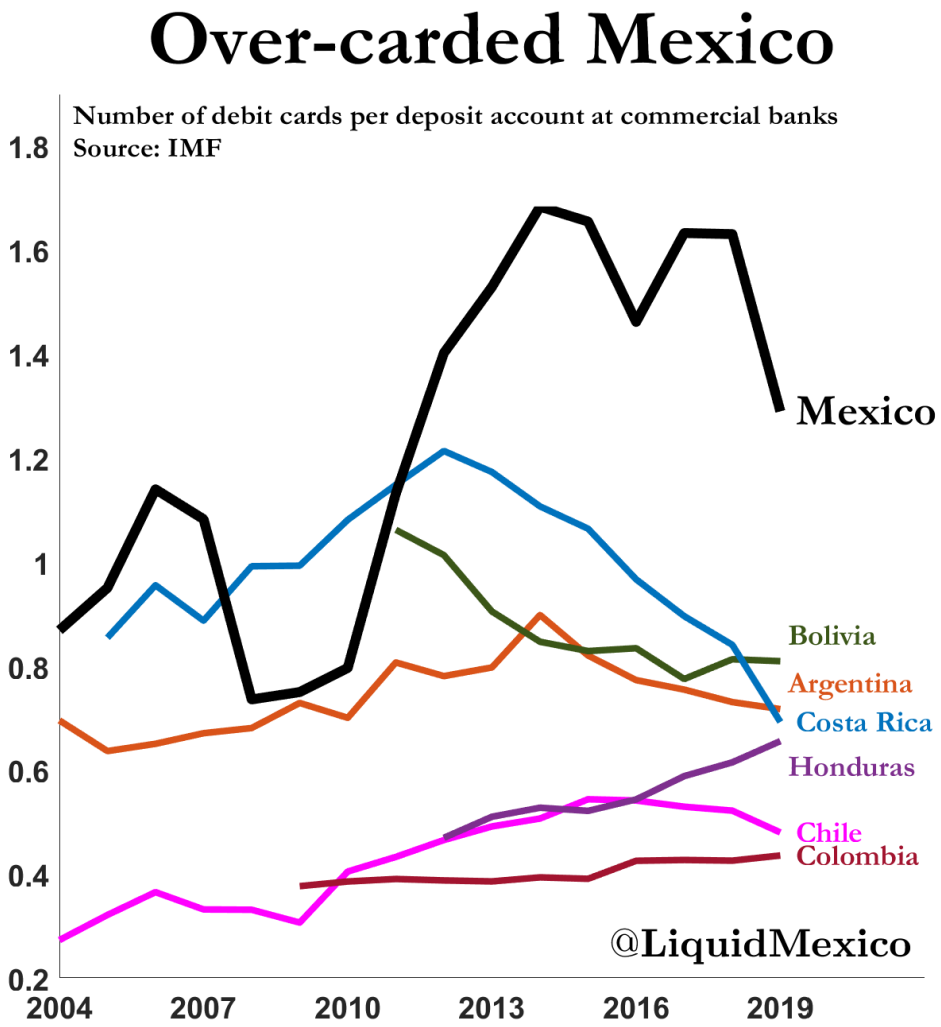
Another revision



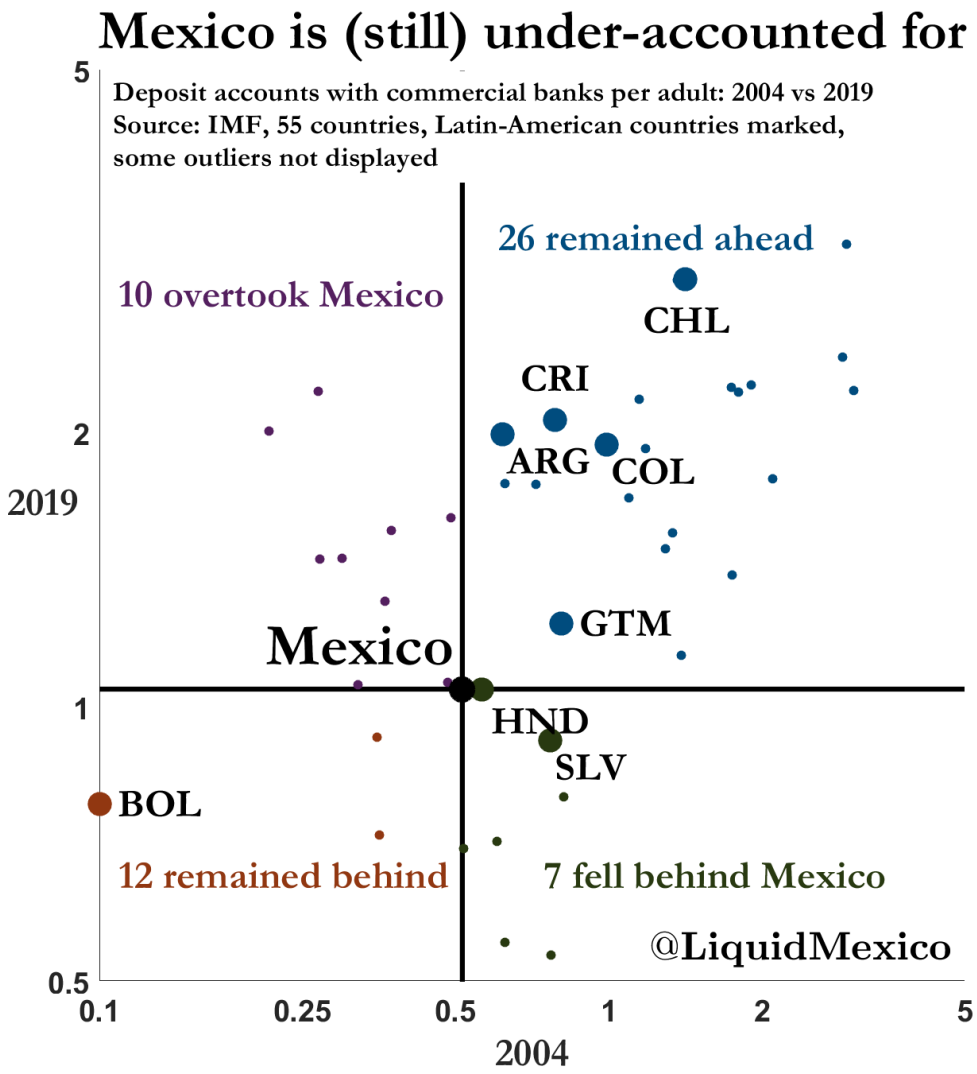
Yet another attempt



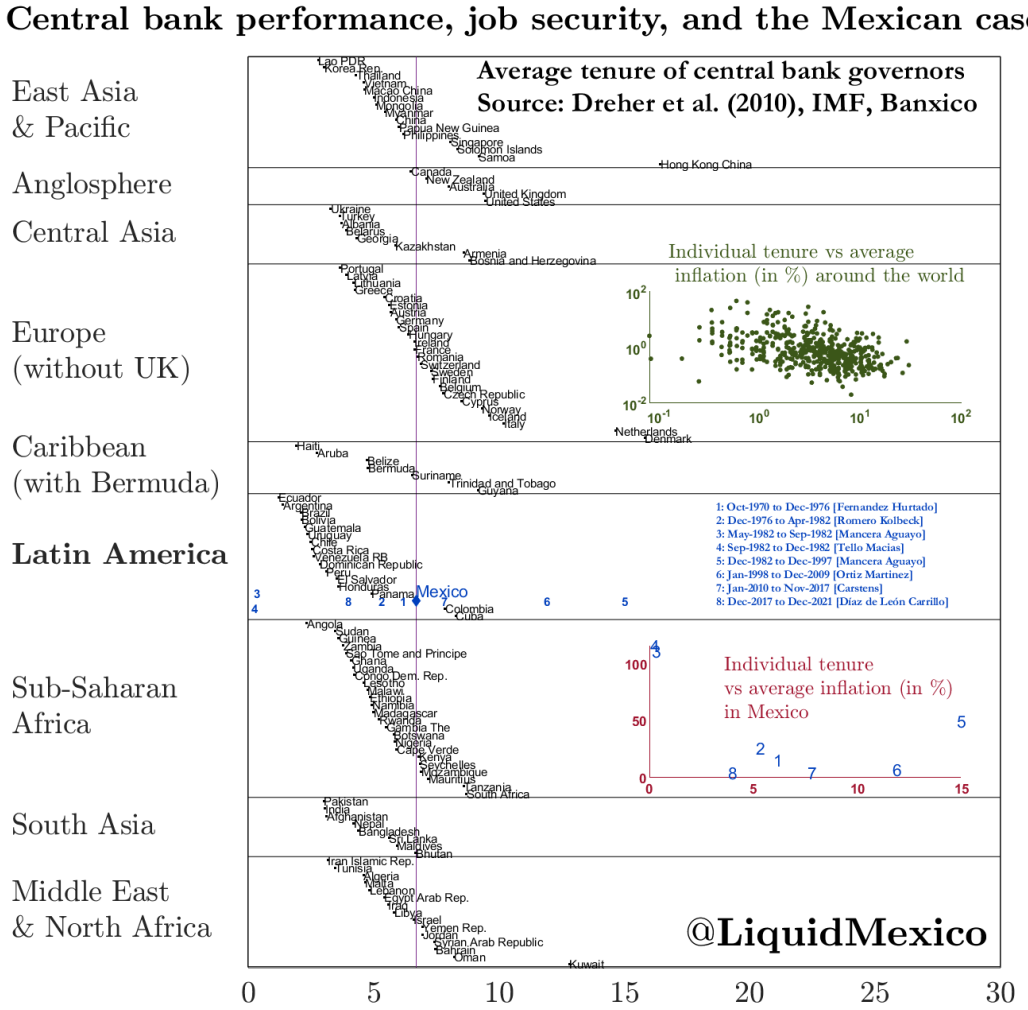
Yet another revision



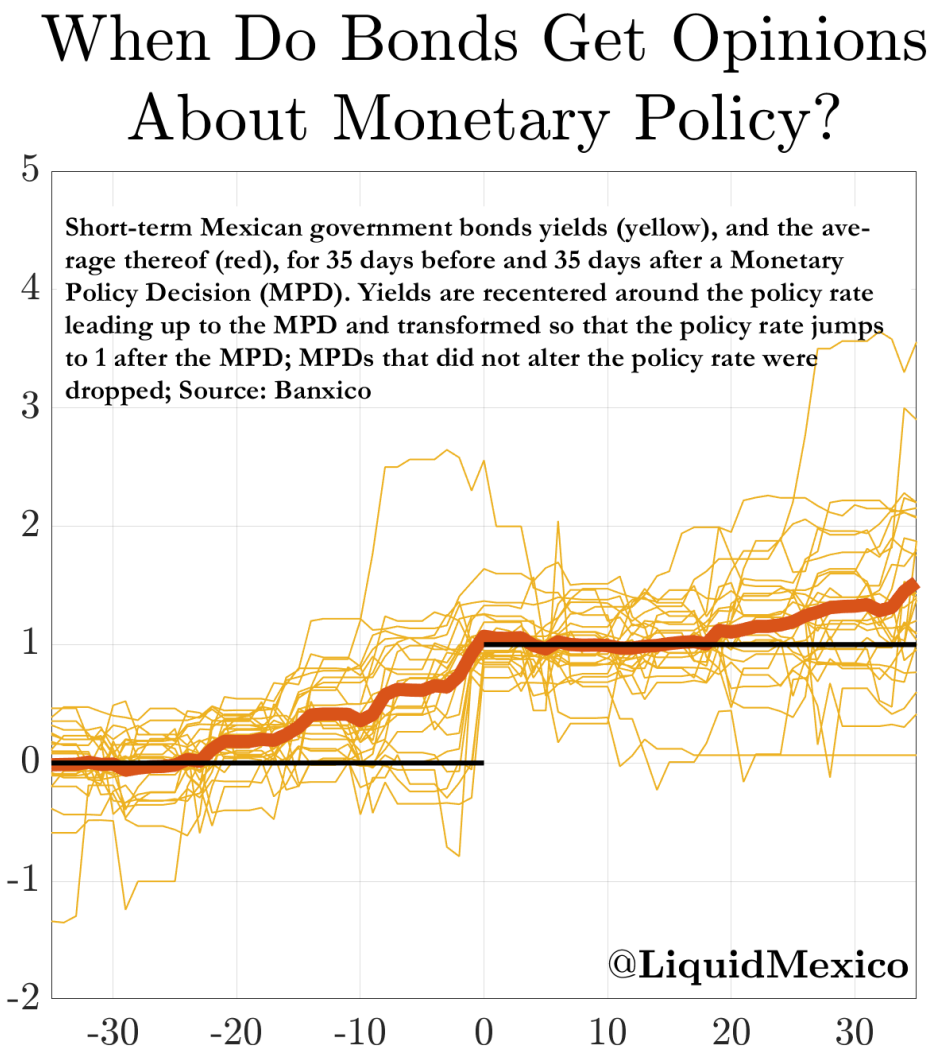
How about this one?



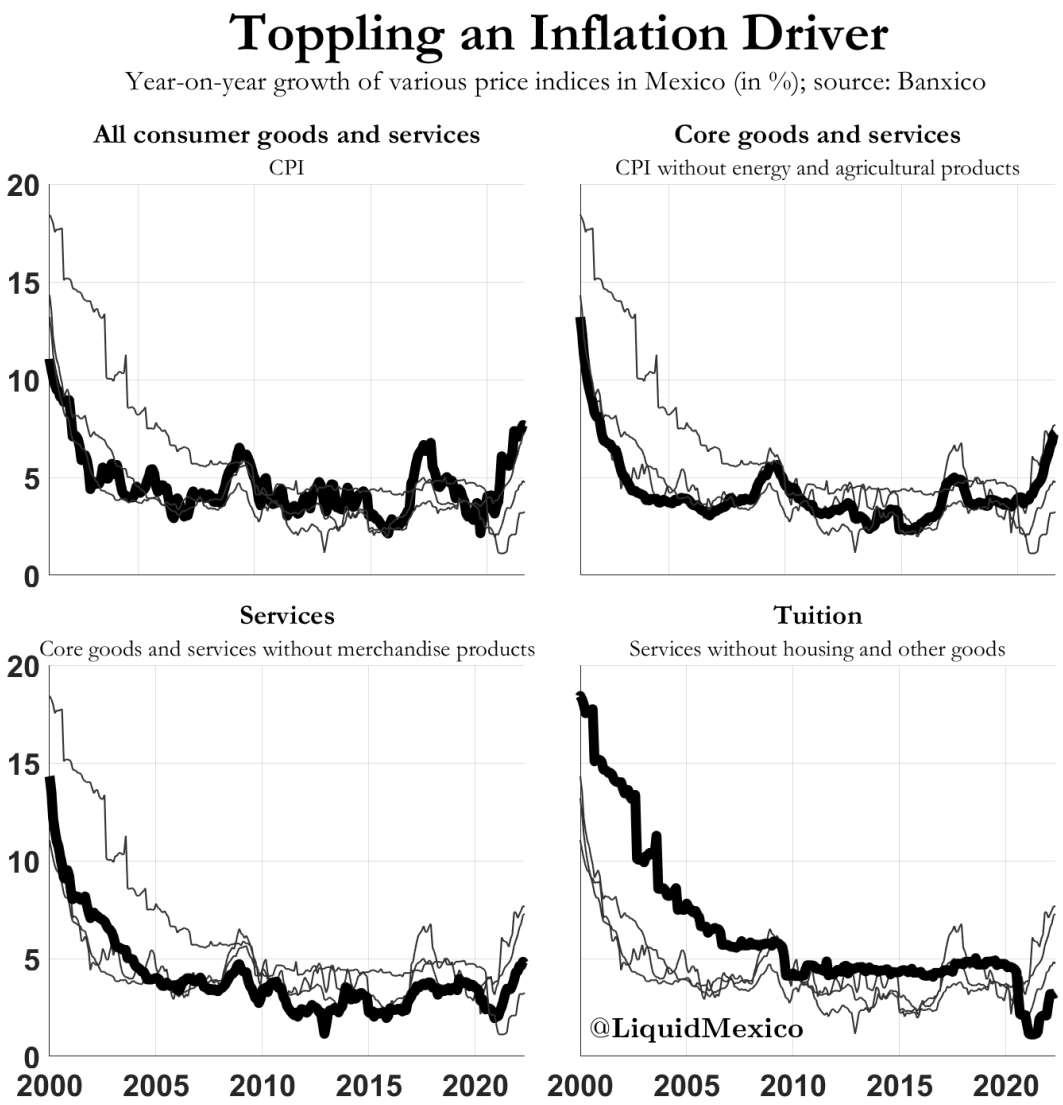
How about this one?



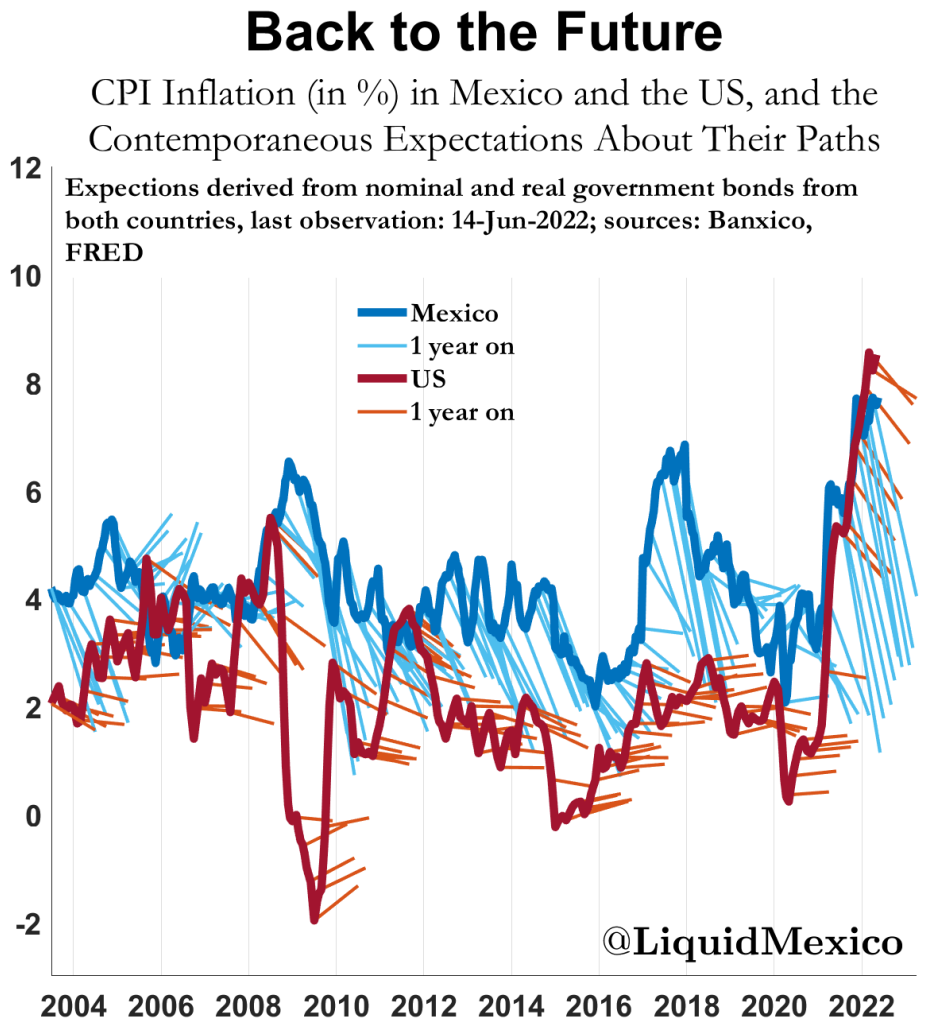
How about this one?



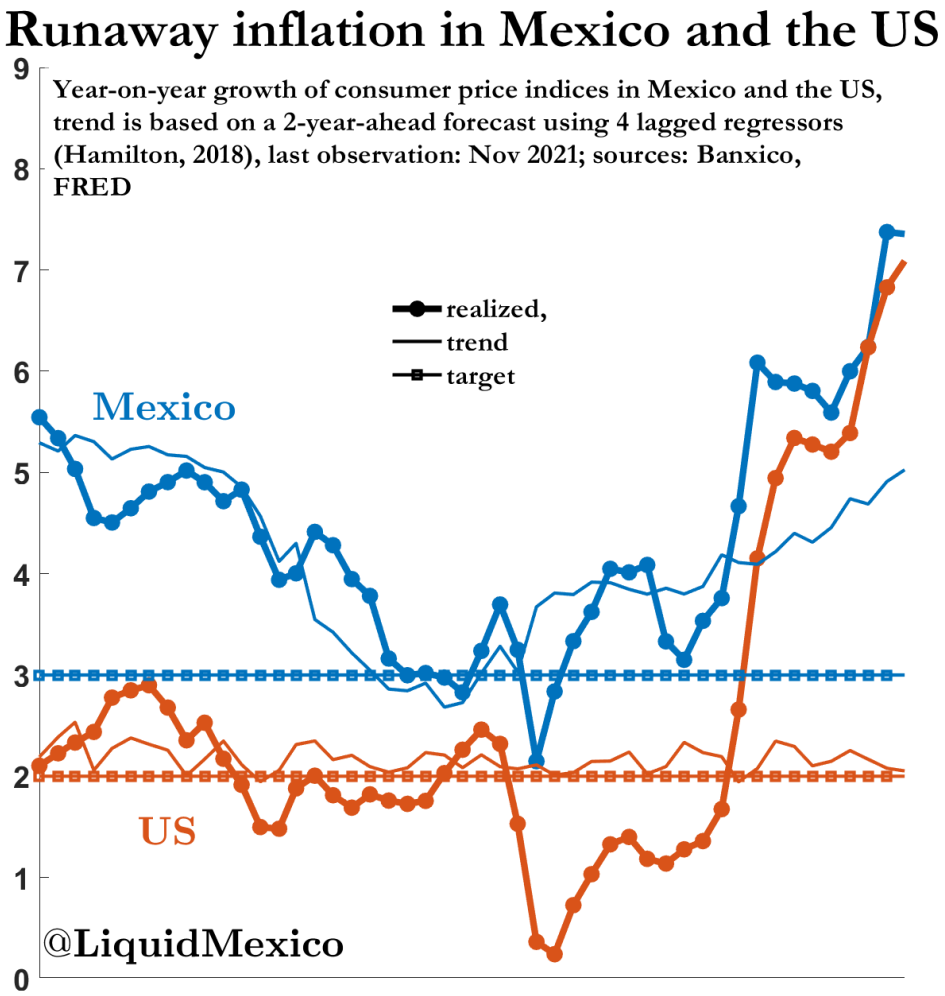
How about this one?



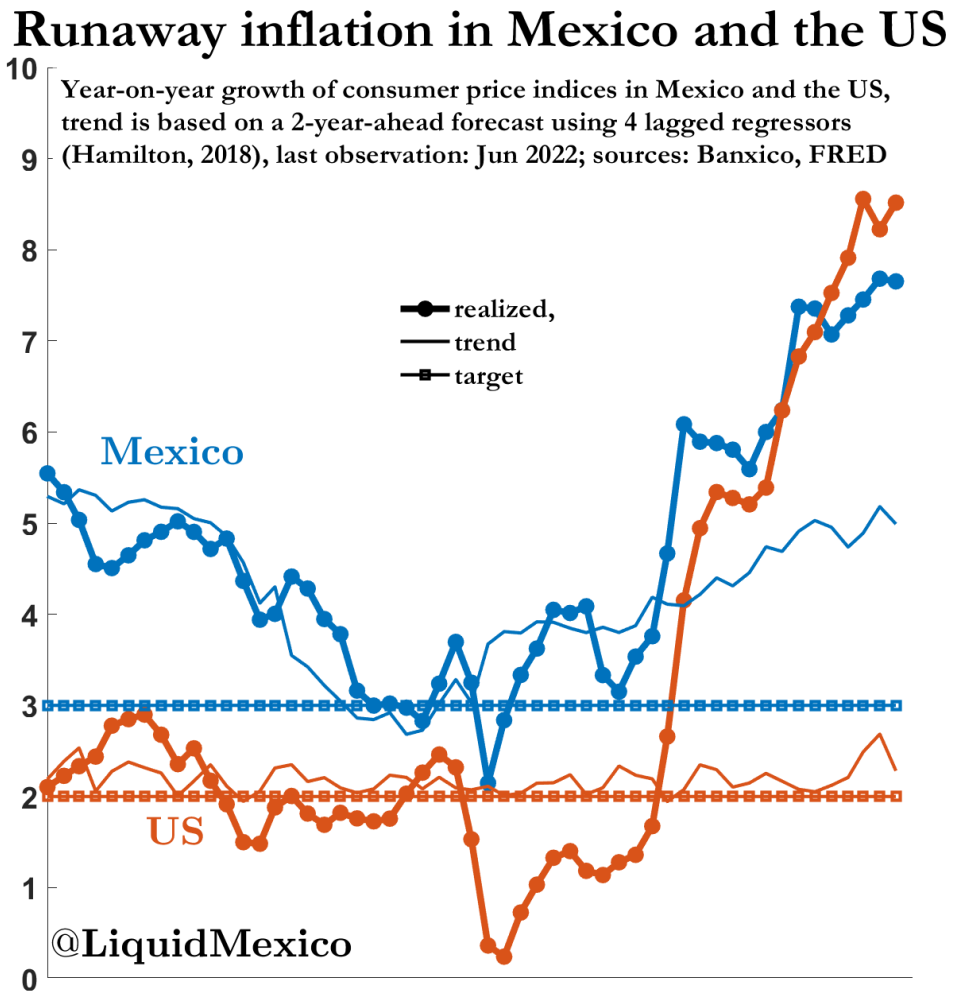
How about this one?



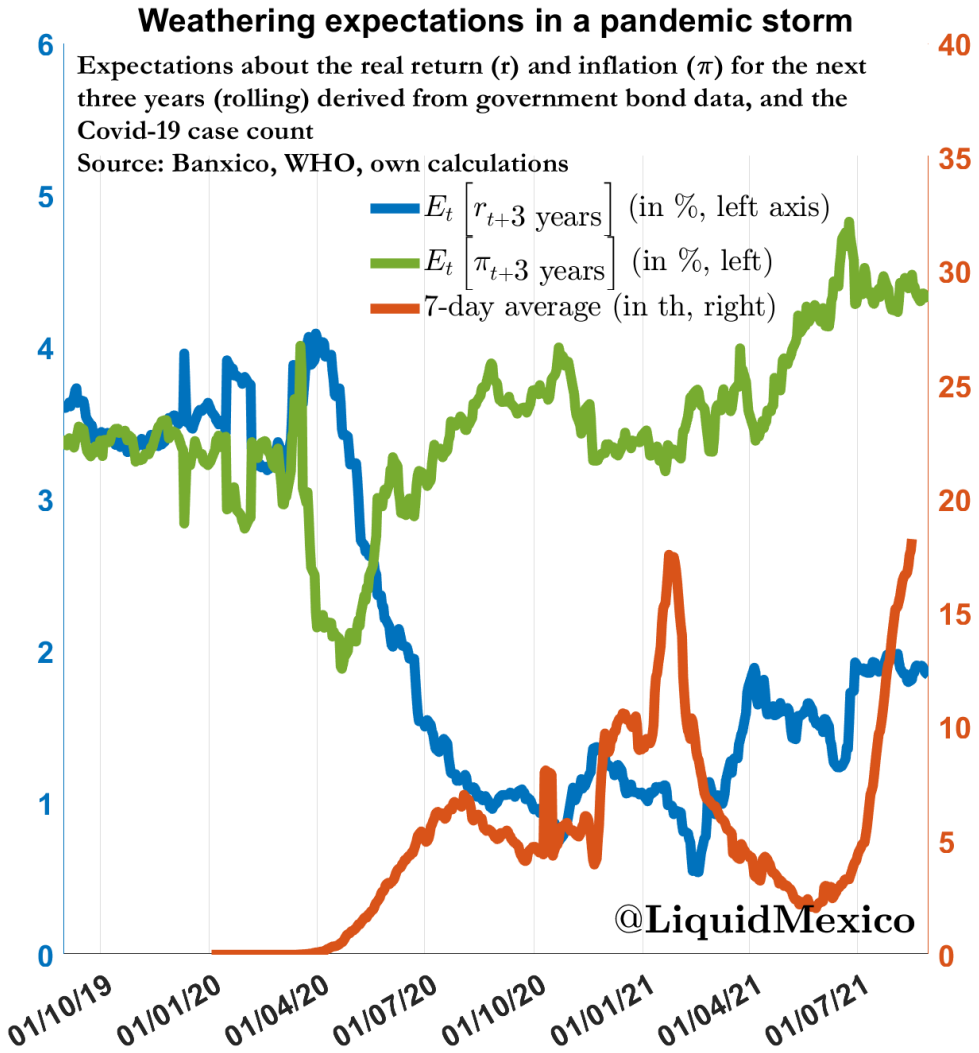
How about this one?



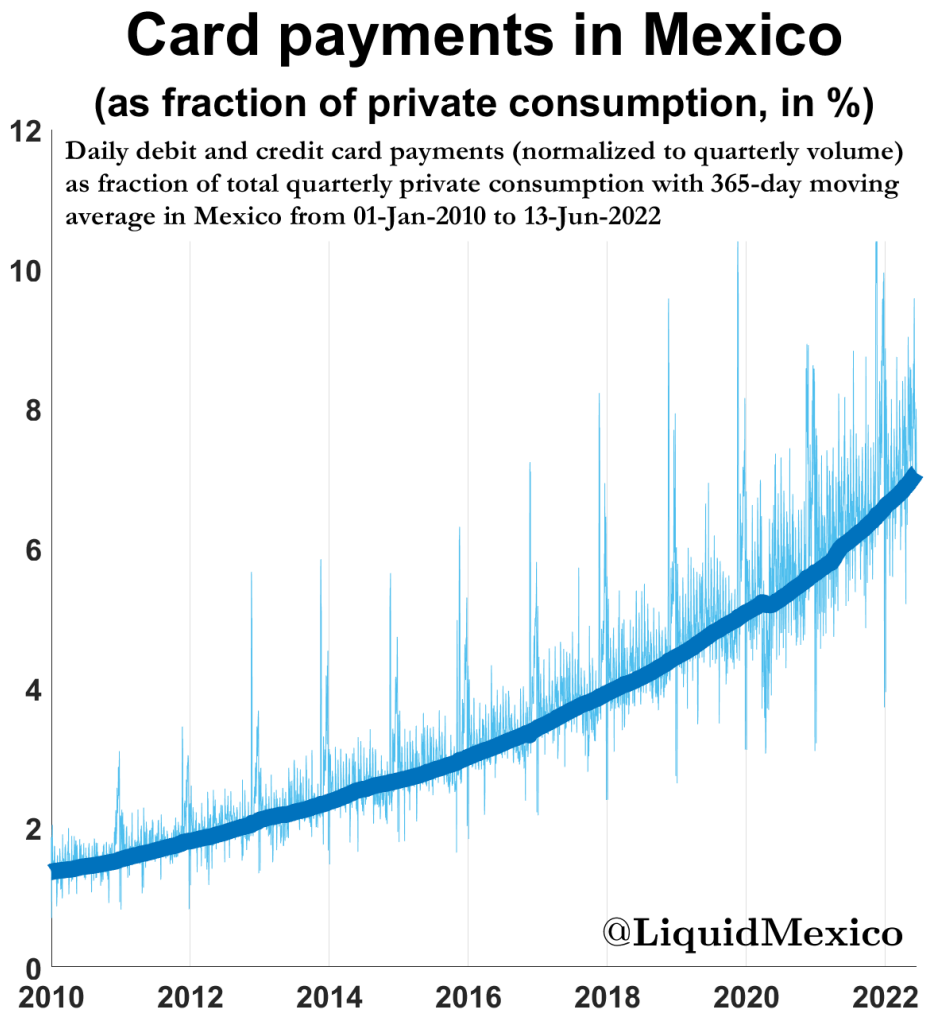
How about this one?



How about this one?



How about this one?



Live data visualization

Monitoring Mexico