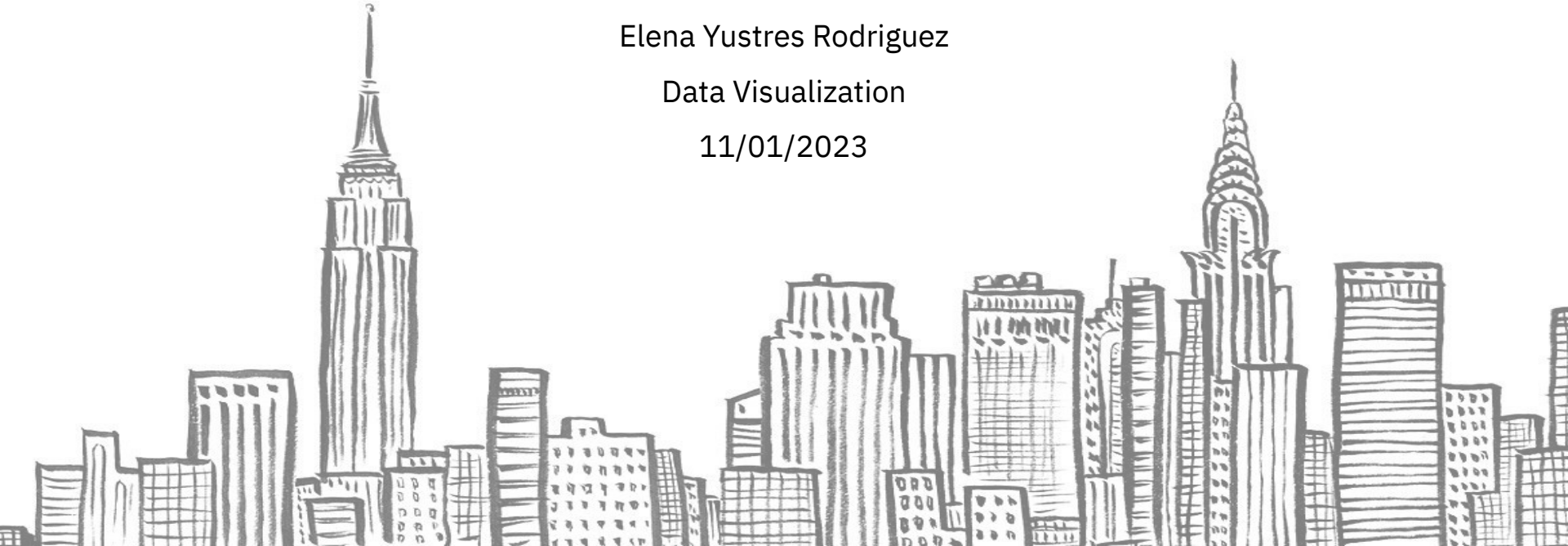
Visualizing the Growth Potential of Chinese Cities: Construction of Maps using *ggplot()* in R

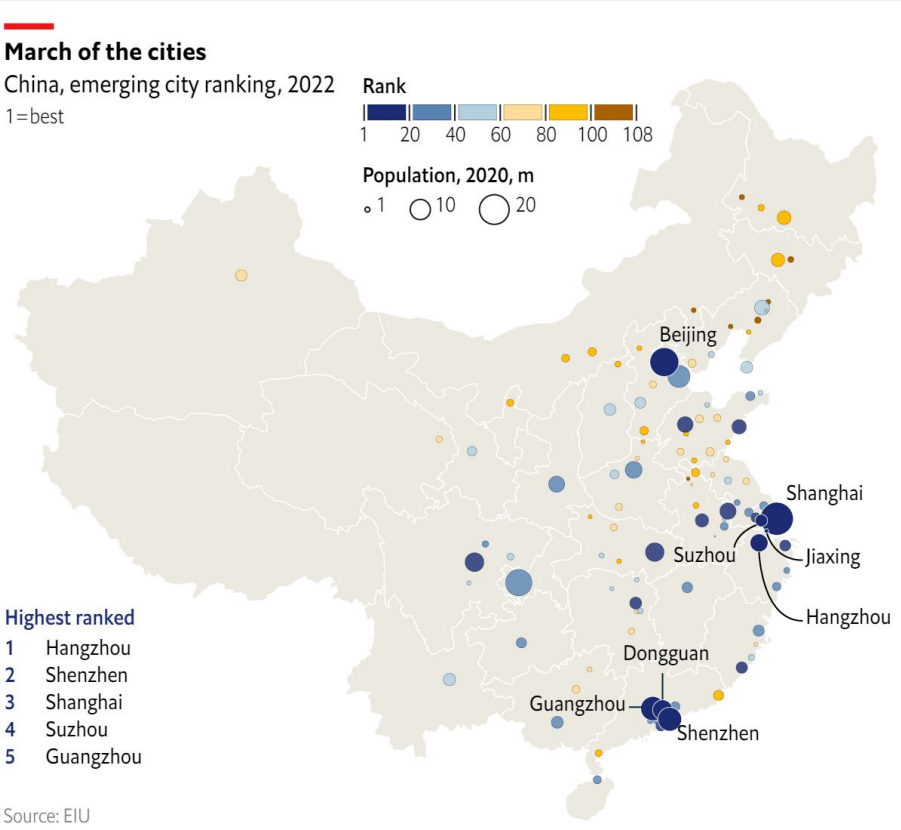
Elena Yustres Rodriguez

Data Visualization

11/01/2023



I. The Original Chart



The Economist

ECONOMIST INTELLIGENCE | EIU

Chart in 'The Economist': "China's GDP is flagging. Where might growth come from?"

- Growth → urbanization → more growth in cities, but which ones?
- Ranking of 106 cities according to comprehensive growth potential scores in "China Emerging City Rankings (2022)" (Economist Intelligence Unit)

Data item: city	*Visual mark*: point
Data attributes: **rank**, **population** and **location**	*Visual channels*: **hue**, **size** and **x-y position**

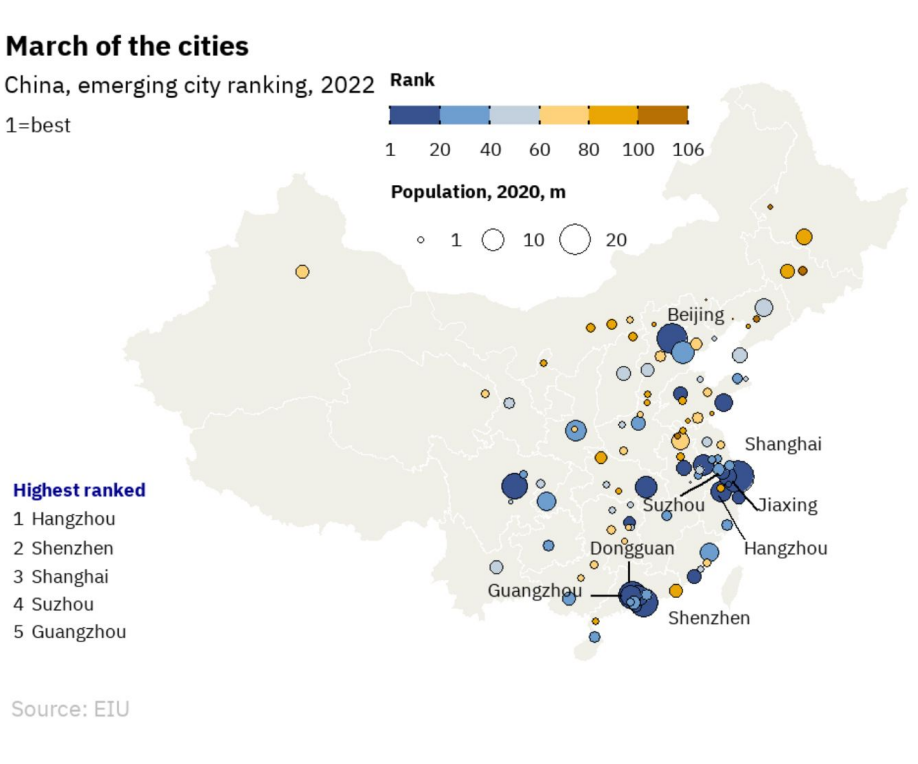
STRENGTHS

- Informative title, subtitle and annotations
- Clear encoding of ranking, population ("area is tricky") and location

WEAKNESSES

- Tufte's data-ink ratio
- Fill legend: choice of color palette and too many bins
- Redundancy of information in annotations of cities

II. Replication: The Highlights

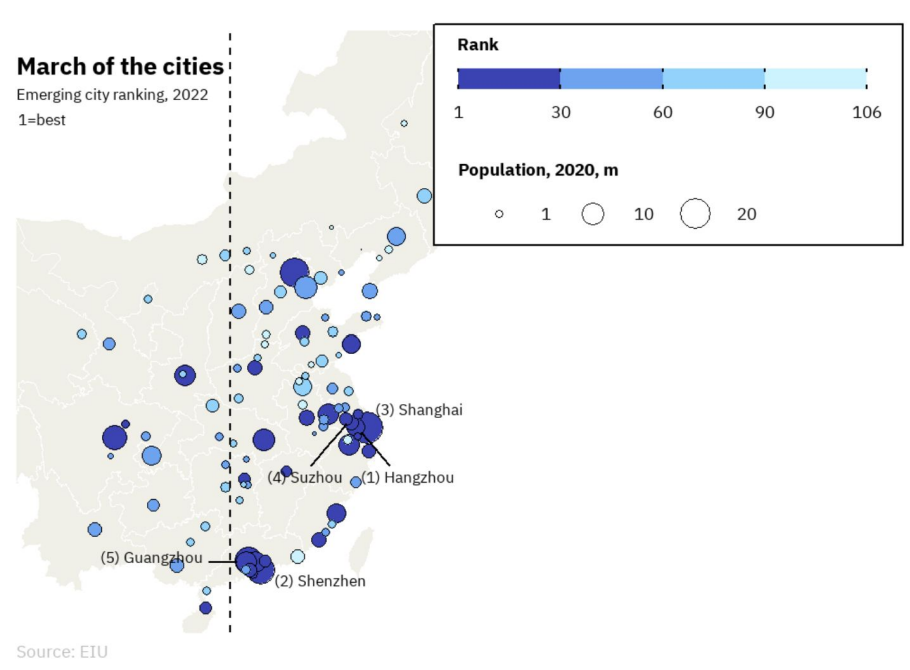


- Collection and processing of the data
  - Four different data sources
  - Missing data
  - Different cities with the same name in different provinces
- Map **aes()** into `geom_polygon()` and `geom_path()` separately
- Scales
  - Working with two different legends:
    - For size, scale_size ≠ scale_size_binned()
    - For color, guide_colorsteps ≠ guide_colorbar
- Annotations

```
p4 <- p3 +
  annotate("text", x=67, y=51, label="1=best",
           size=6, family="IBM Plex Sans") +
  annotate("text", x=65, y=30, label="Highest ranked",
           size=5, family="IBM Plex Sans", hjust=0, vjust=0, angle=0,
           fontface="bold", color="blue4") +
  ...
  annotate("text", x=122, y=26, label="Hangzhou",
           size=5, hjust=0, vjust=0, family="IBM Plex Sans") +
  annotate("text", x=116, y=21, label="Shenzhen",
           size=5, hjust=0, vjust=0, family="IBM Plex Sans") +
  ...
  annotate("segment", x = 113, y = 25.5,
           xend = 113, yend = 24) +
```

!

III. Potential Enhancements



(1) Maximize data-ink ratio: define coordinates and "zoom in"

p + coord_map(xlim=c(100, 155), ylim=c(20, 50))

(2) Dashed line to show that only one city among "top 10" is west of 112°

geom_vline(xintercept=112, linetype="dashed")

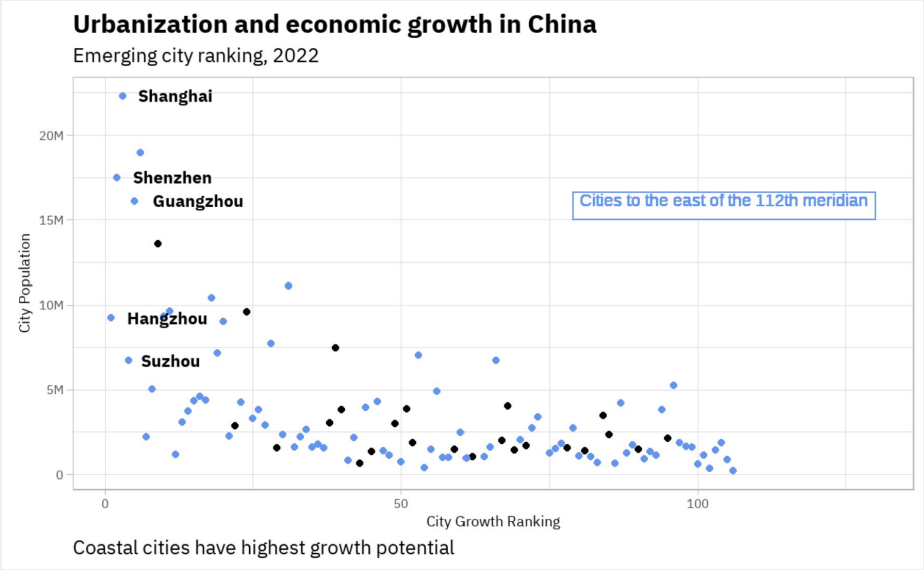
(3) Fill scale:

(a) Change of color palette: darker shade = higher growth potential

(b) For discriminability, 6 → 4 bins

(4) Simplification of annotations: label **and** ranking on the map

IV. Alternative Visualization



- Key idea of the chart: convey the relationship (*if there is one*) between growth potential, population and location
- Hence:
  - Population vs. ranking: negative relationship, but higher spread for highest-ranked cities
  - Location: unclear relationship with ranking, but clear majority of cities in the dataset are on the Coast
- Conclusion: the coast is the engine of China's development

Q & A

Thank you for your attention!