Lecture 3
Data Visualization

Pierre Biscaye

Université Clermont Auvergne

Data Science for Economics

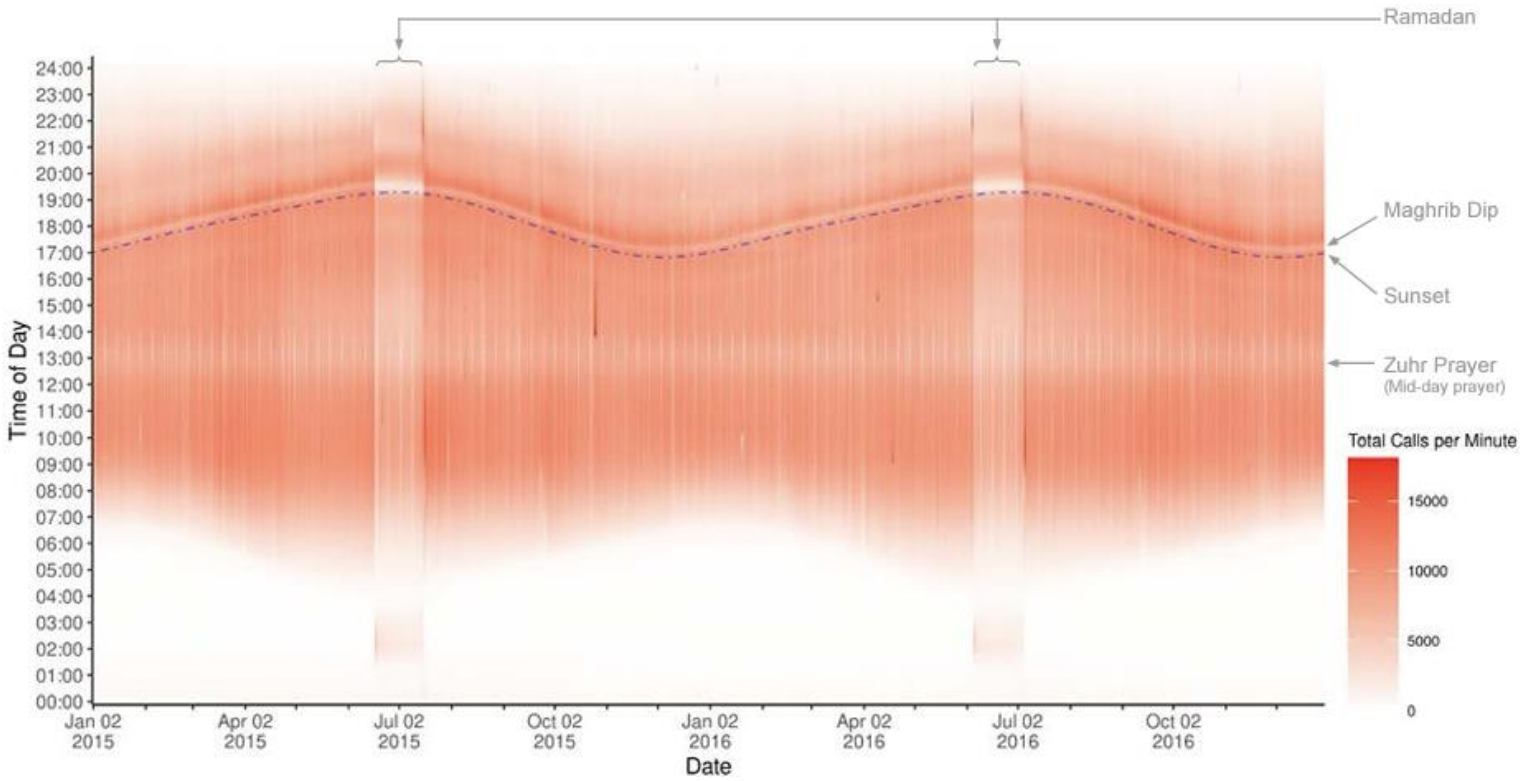
Agenda

1. Value of visualizing data
2. Visualization examples in 2024 dev econ Job Market Papers
3. Intro to data wrangling and visualization in Python

A picture is worth a thousand words

The best research can be summarized in one key exhibit.

Figure 2: Intensity of mobile phone calls over time



Dube et al (2022 WP), Measuring religion from behavior: Climate shocks and religious adherence in Afghanistan

Nice examples of spatial data visualization and development: Jose Luengo-Cabrera

Value of visualizing data

- **Simplifies Complex Data**: Transforms multidimensional datasets into easily interpretable visuals, enabling researchers to identify patterns, trends, and outliers at a glance.
- **Enhances Communication**: Effectively conveys findings to diverse audiences.
- **Facilitates Hypothesis Generation**: Helps uncover unexpected relationships or anomalies in data, prompting new.
- **Supports Robust Analysis**: Enables exploratory data analysis, model validation, and assumption testing to strength formal econometric analysis.

Many different types of visualizations

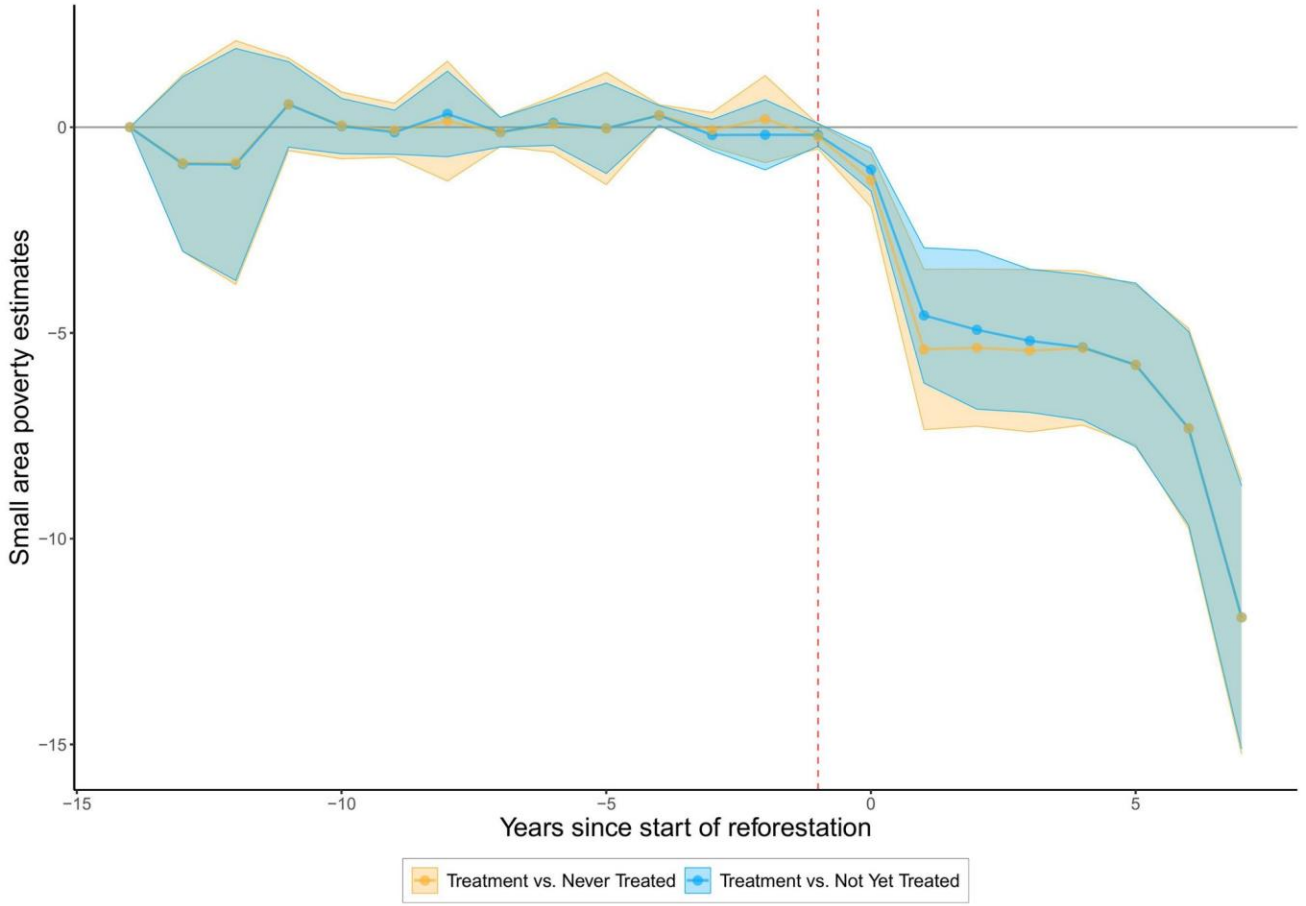
- Frequency visuals: bar charts, histograms, box plots, pie charts, summary tables
- Correlation visuals: scatter plots, regression tables
- Trend visuals: line graphs
- Maps: heatmaps, chloropleth maps
- Network diagrams
- Logical/hierarchical visuals: tree maps, flow charts
- And more…

Source

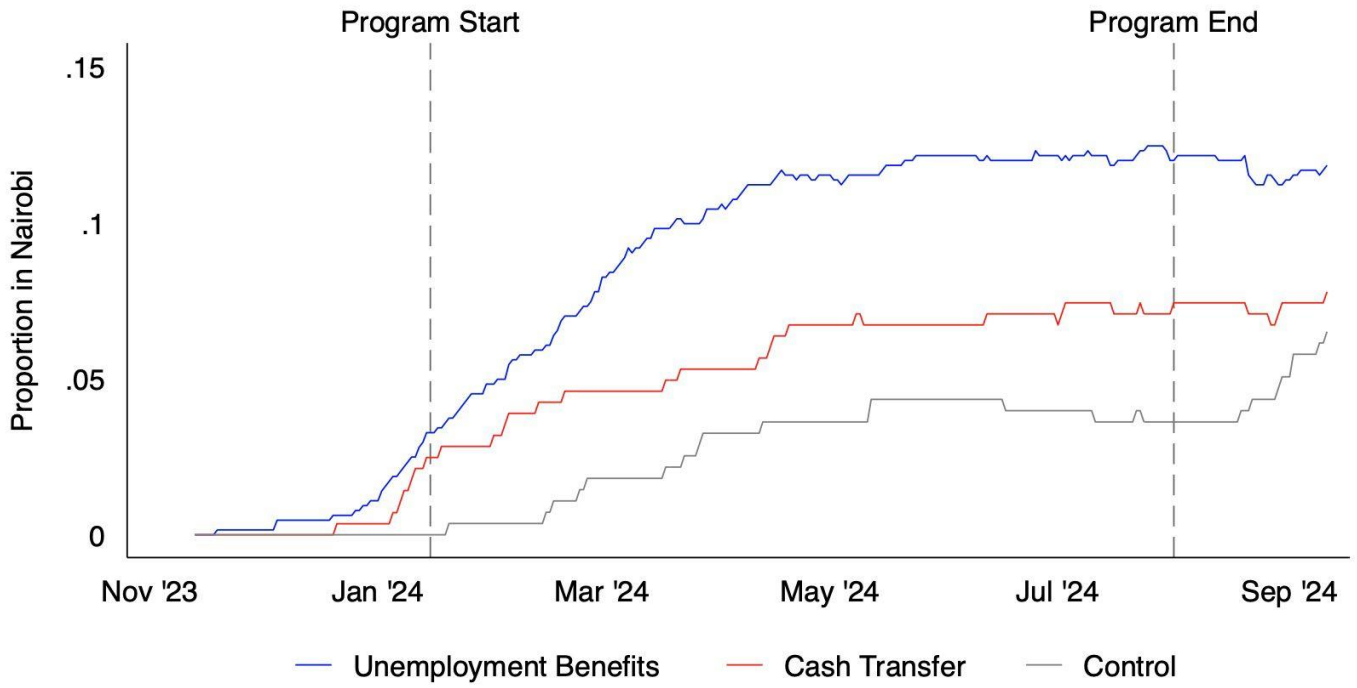
Visuals in papers of current econ dev job market candidates

- Nice examples of what current PhD candidates are working on
- Wide variety of methods and visuals
- Applications to economic development research
- Main source: 2024 WB JMP Blogs

Pagel: Tree planting (National Greening Program) and poverty in the Philippines

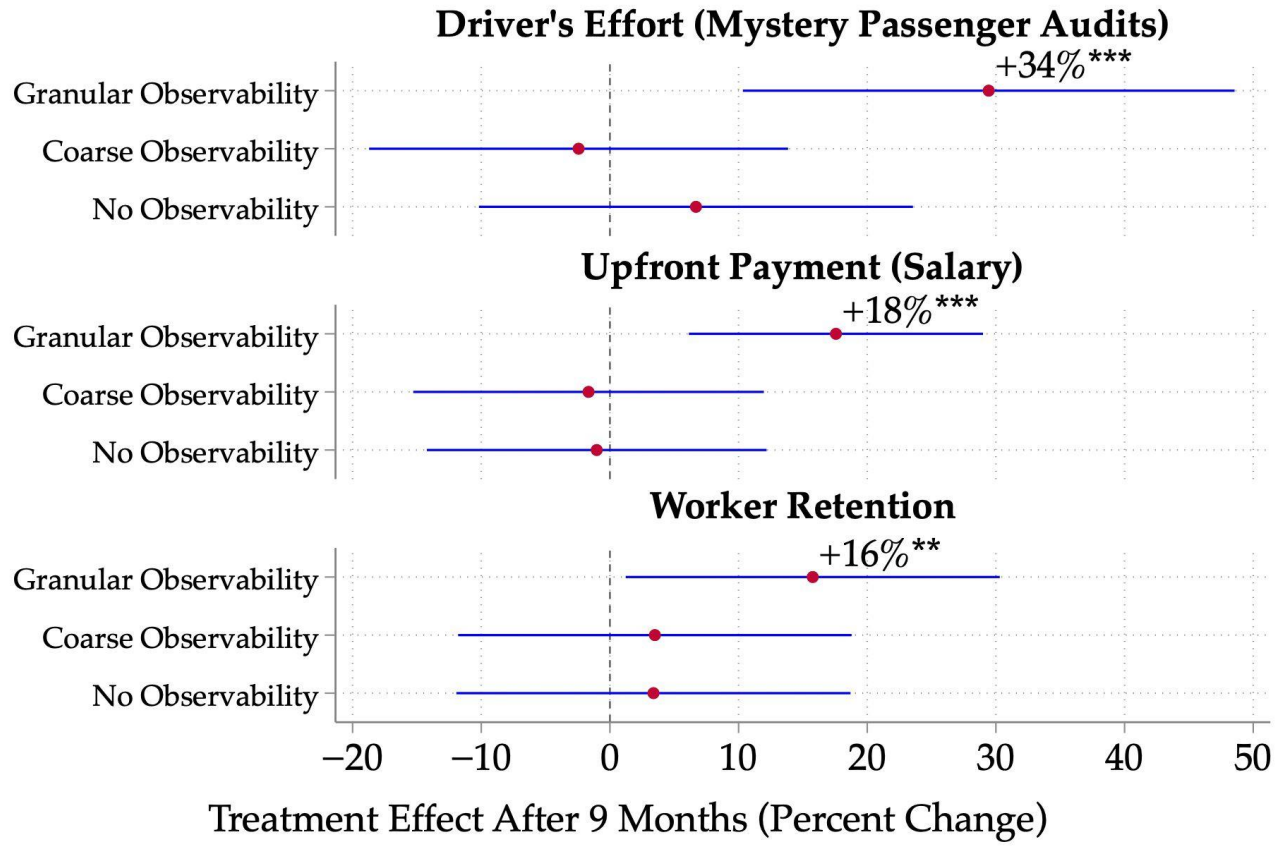


Miner: Reducing internal migration uncertainty RCT in Kenya



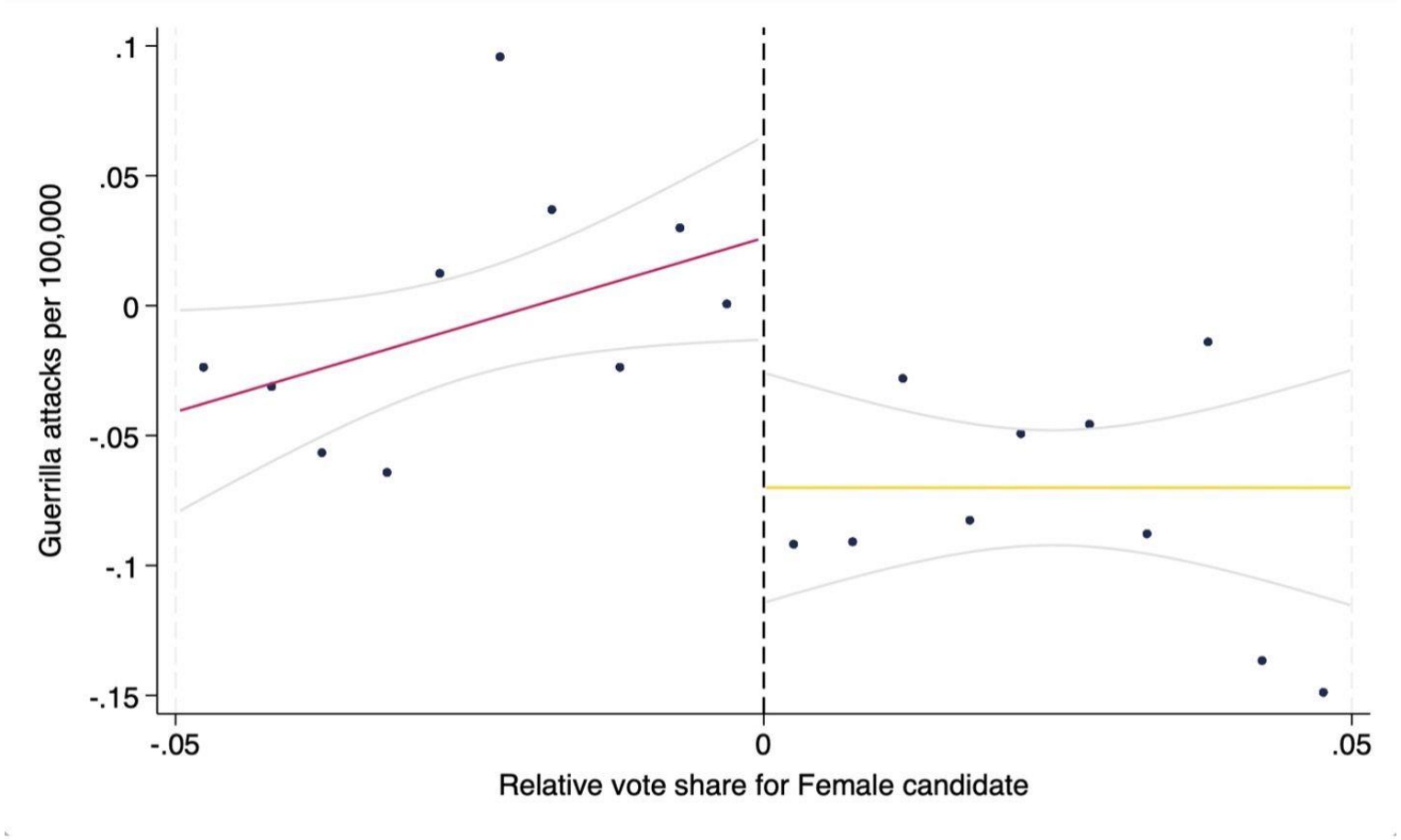
Unemployment benefits conditional on moving to Nairobi
Cash transfer unconditional

Houeix: Digital technology, information frictions, and firm efficiency among taxis in Senegal



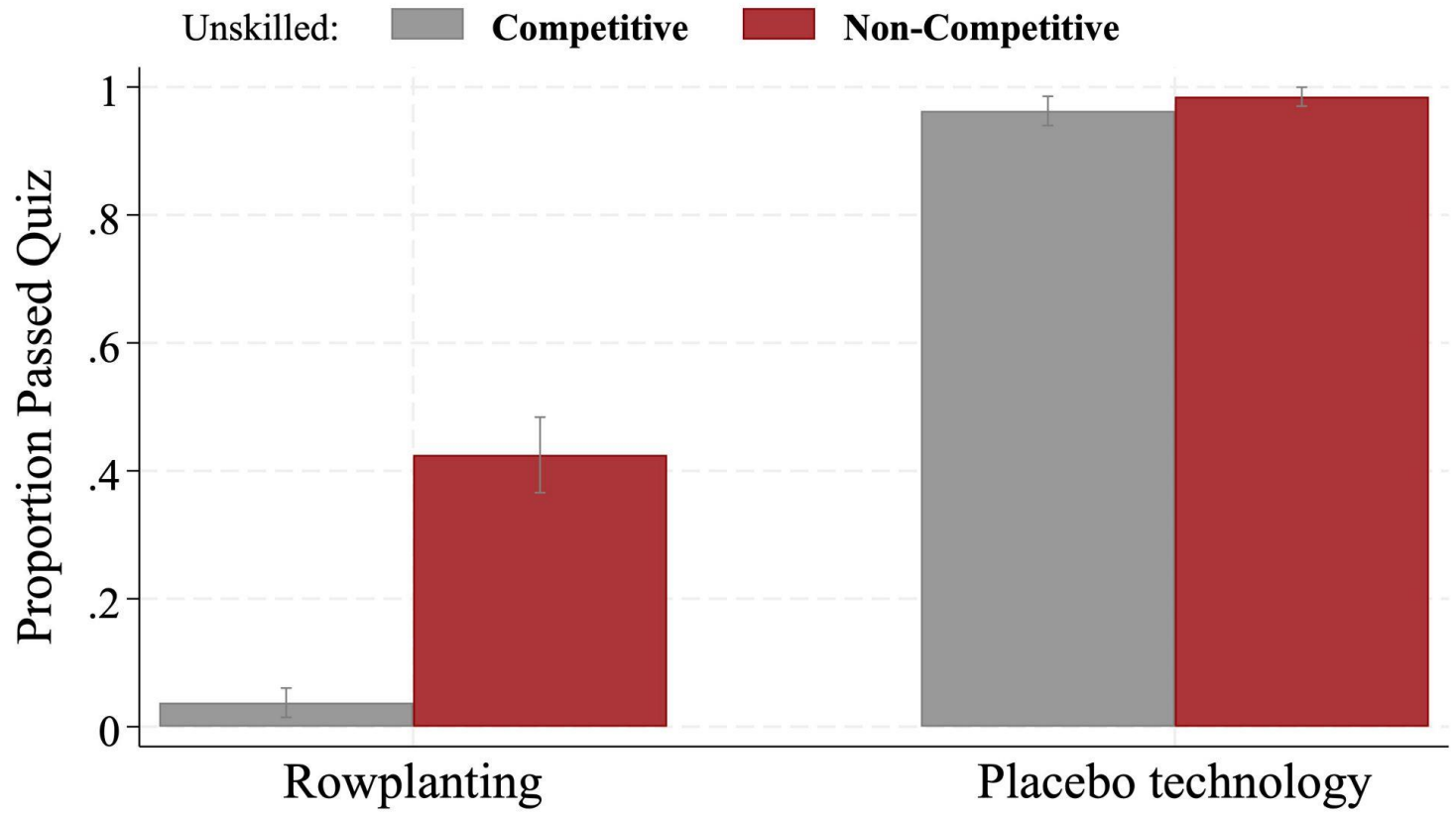
Granular: owner can observe driver's effort and output
Coarse: owner gets noisy measure of output

Eslava: Female leaders and armed conflict in Colombia



Municipality-level data

Cefala: Knowledge hoarding as a barrier to agricultural innovation in Burundi



Competitive = trainer and trainee in same village

Cisse: Value of electricity reliability in Senegal

Figure A3: MAP OF FEEDER LINES BY TIMING OF RELIABILITY PROJECT IN SENEGAL

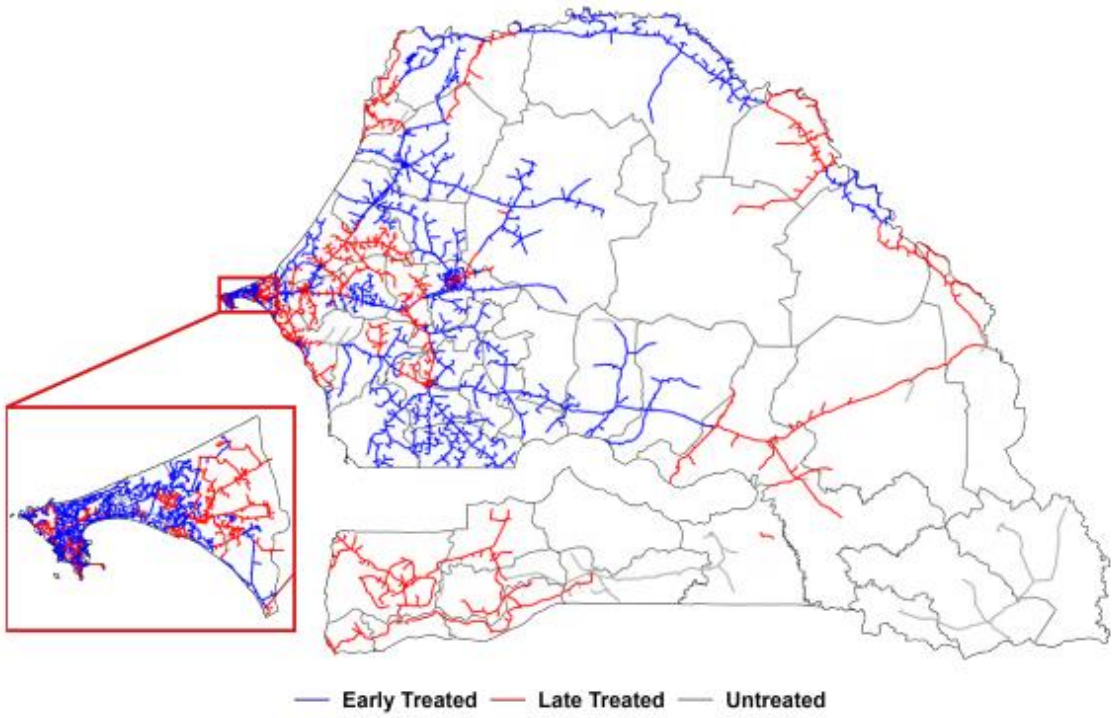
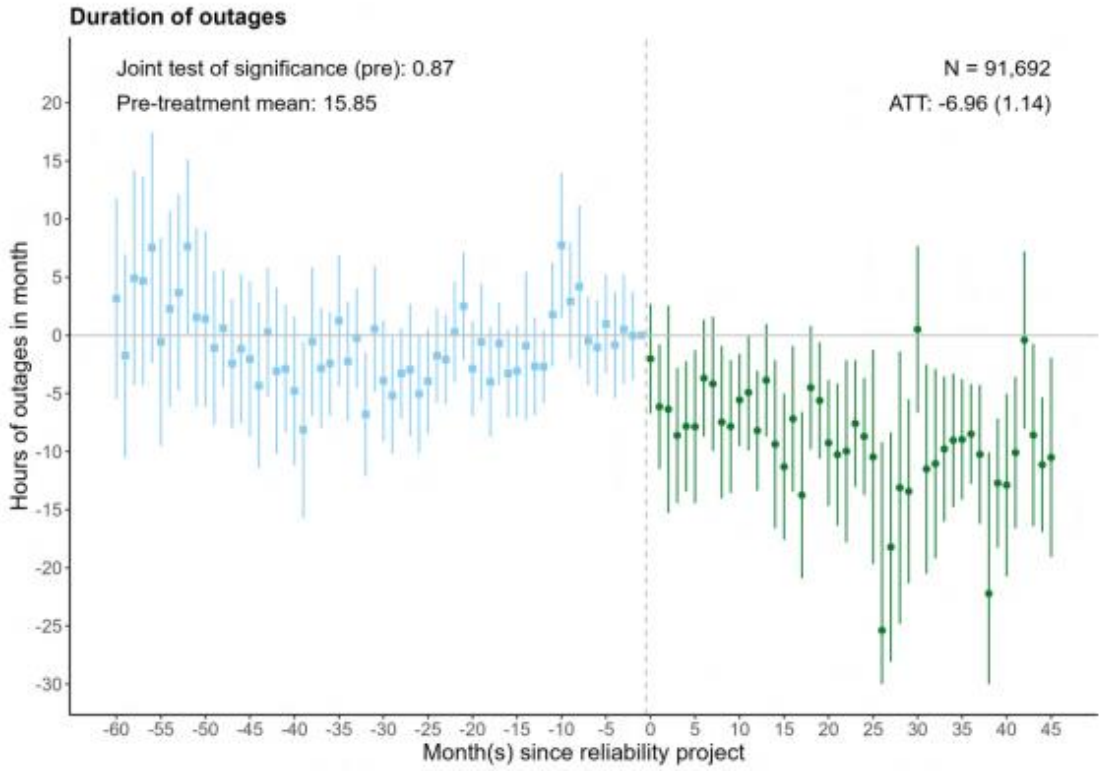
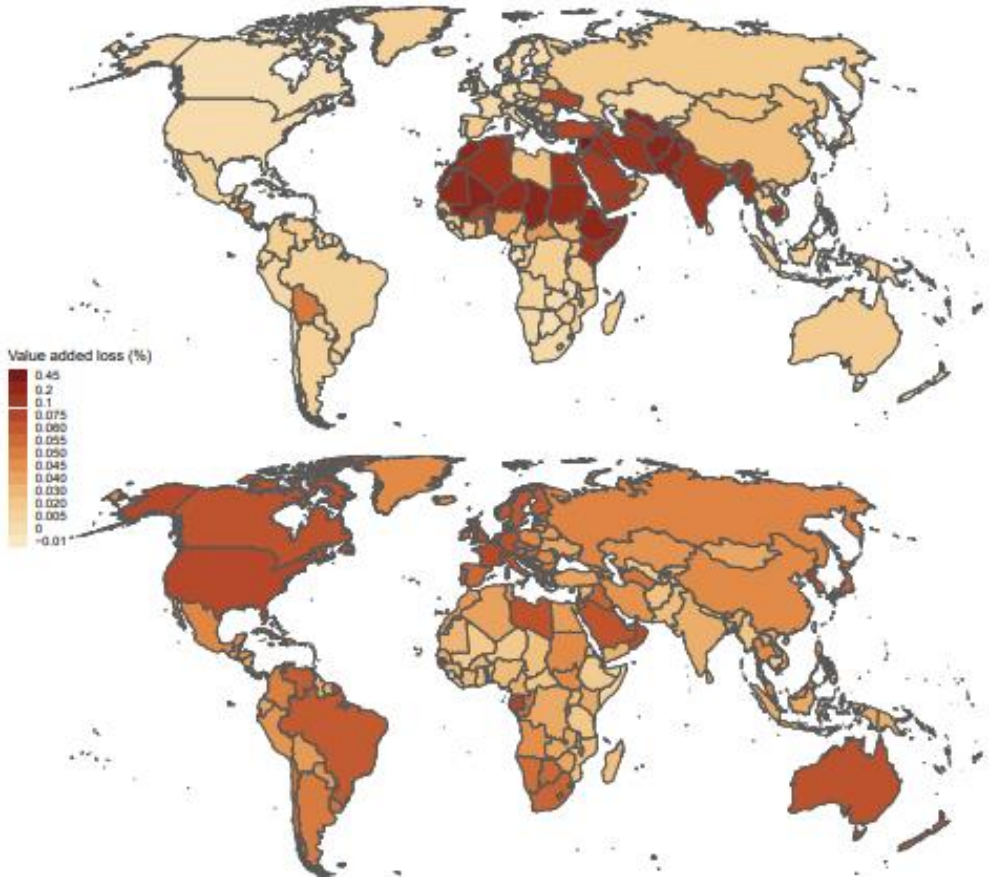


Figure 3: IMPACT OF RELIABILITY PROJECTS ON THE MONTHLY DURATION OF OUTAGES



Zappala: Impacts of extreme heat in agricultural across sectors and space

Figure 8. Annual value added losses (%) induced by recent warming (2001-2020)
(a) Global distribution of value added losses in agriculture (above) and in downstream sectors (below)



Value of visualization along the research process

- Data cleaning and preparation:
  - Summary statistics and frequency distribution
  - Checking for outliers
  - Identifying potential data errors
  - Thinking about data generating process
- Testing assumptions: plotting correlations, simple regressions
- Econometric analysis: synthesizing and clearly communicating results

Think about your audience

- You don’t write an academic paper the same way as a policy brief
- Your visualizations should be at the level of the audience

Into Jupyter!