

Data Visualization

Kristo Raun

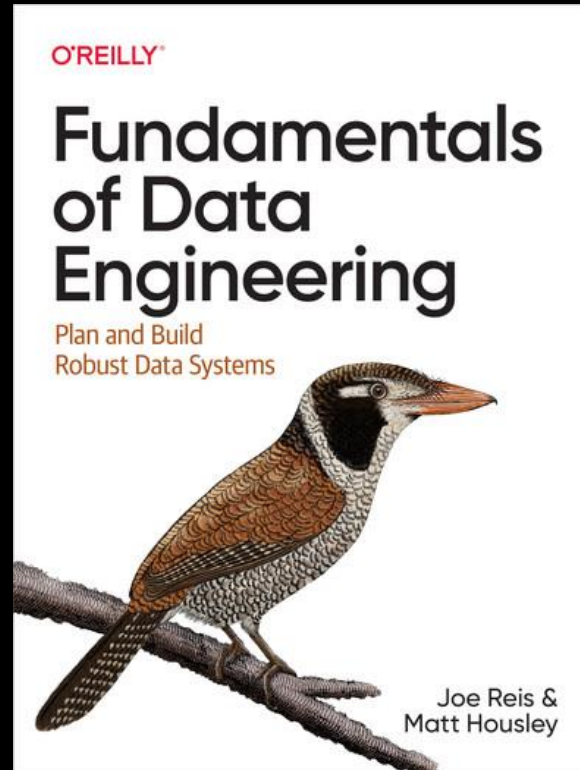
Data Engineering 2024 Fall



UNIVERSITY OF TARTU

Reading

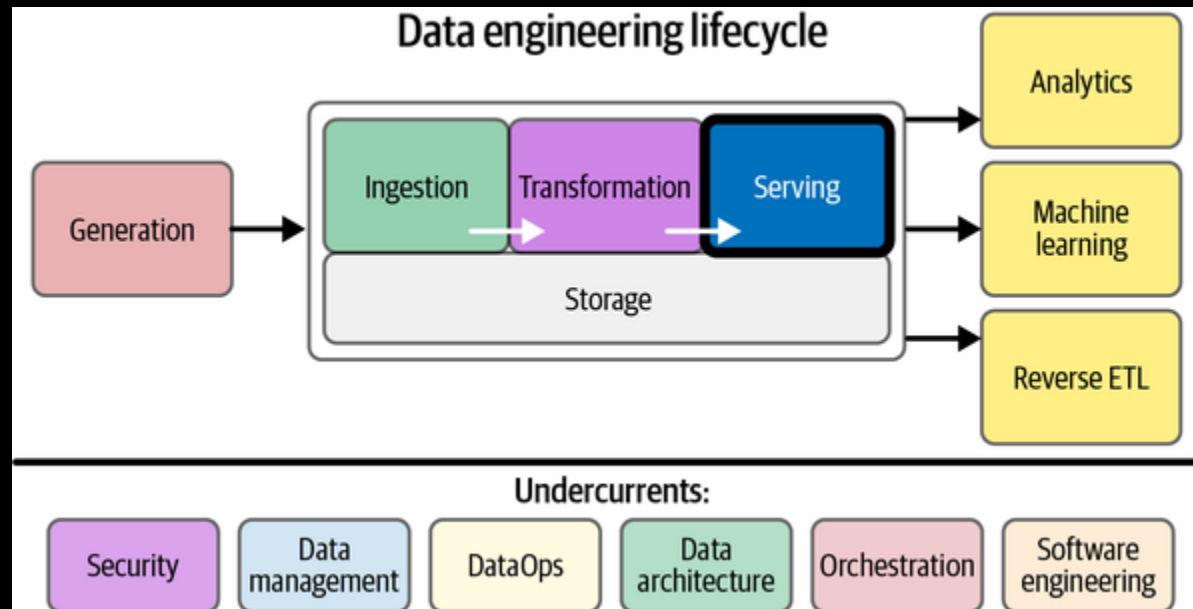
- Chapter IX



Agenda

- Serving Data
- Data Visualization
- Quiz session

Serving data



Serving data

- Serving data for downstream use cases:
 - BI & analytics
 - Reporting, dashboards
 - Machine learning
 - Data for ML pipelines (model training and inference)
 - Reverse ETL
 - Back to source (OLTP) systems

Serving data

- How to enable data consumers access to data?
 - SQL access
 - Most common and flexible
 - Requires consumers to have SQL proficiency
 - API (REST)
 - High development and maintenance cost
 - More control
 - Export to files
 - Static
 - Generally easy regardless of end user/system

Serving data - considerations

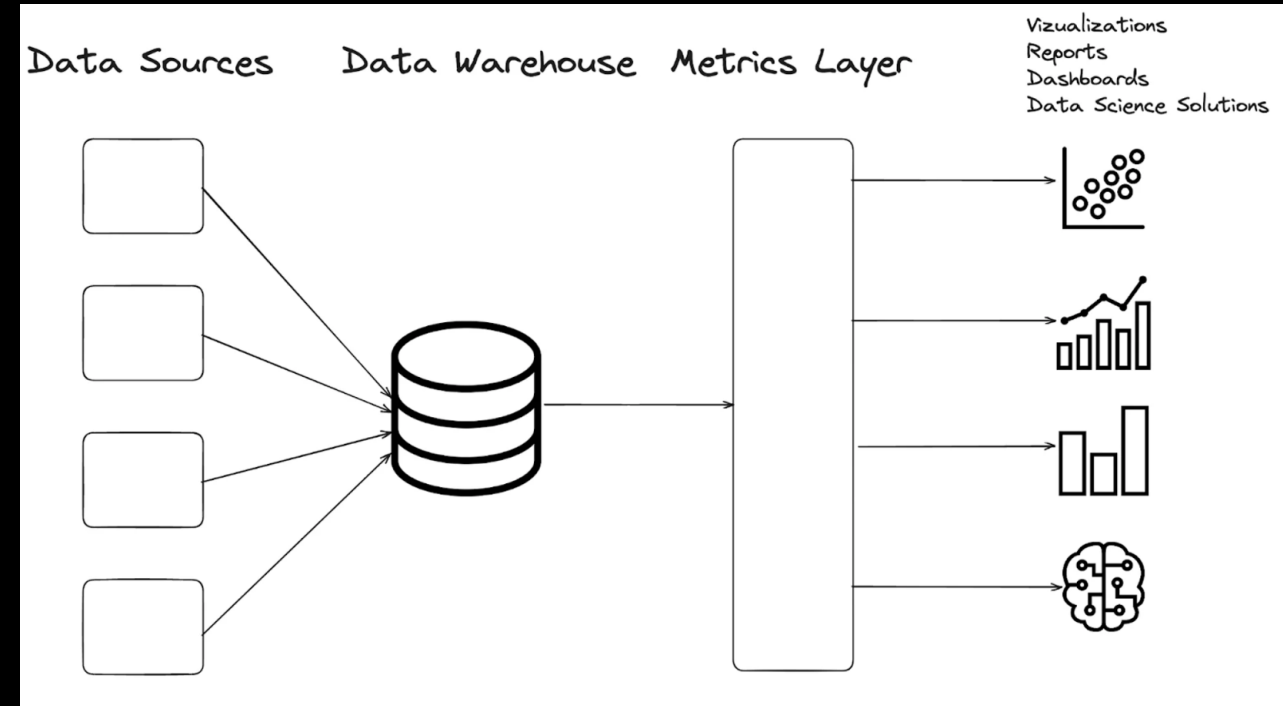
- Data latency
 - Views vs tables
 - Pre-aggregated tables
- Security & Privacy
 - Access
 - Monitoring
- Governance
 - Metadata, catalogs

Serving data – “trends”

- Self-service BI
 - Data usage by non-technical users
 - Critique: use SQL
- Data mesh
 - Domain-oriented data products
 - Critique: complexity and manageability

Serving data – Metrics layer

- Motivation
 - Why is the value different?
 - Meaning of values?
- Same metrics for all consumers
- Similar concept: semantic layer
 - What's “revenue”?



Agenda

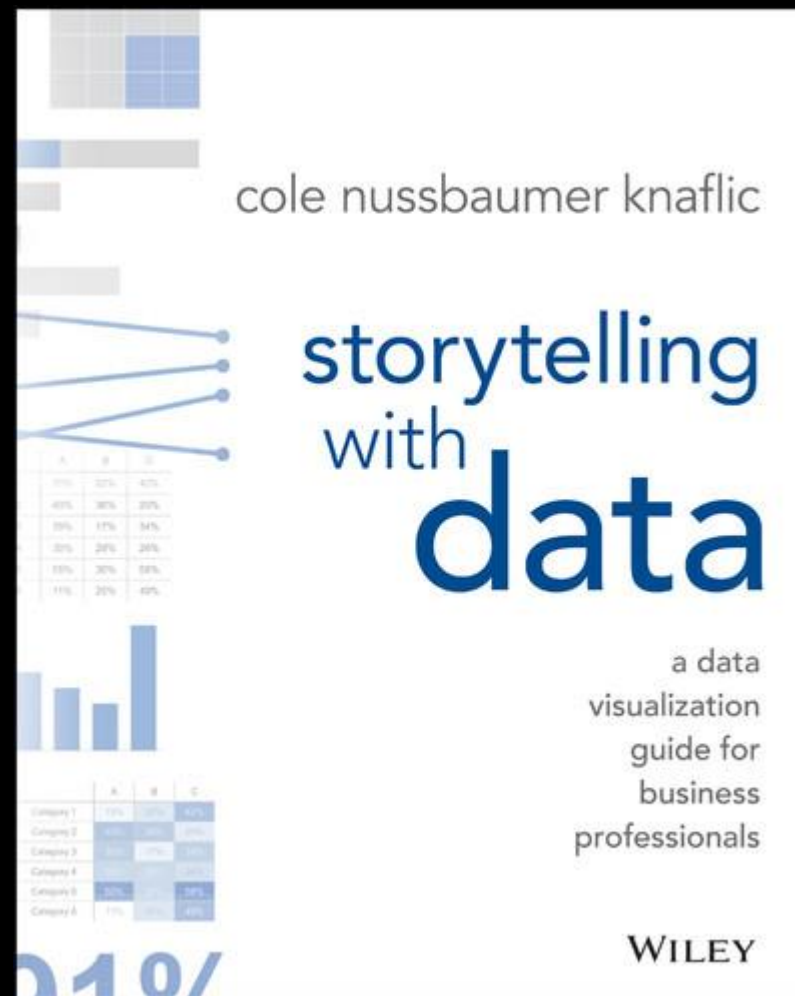
- ~~• Serving Data~~
- Data Visualization
- Quiz session

FIRST, LET'S RECAP The IMPORTANCE of CONTEXT

TYPES of ANALYSIS



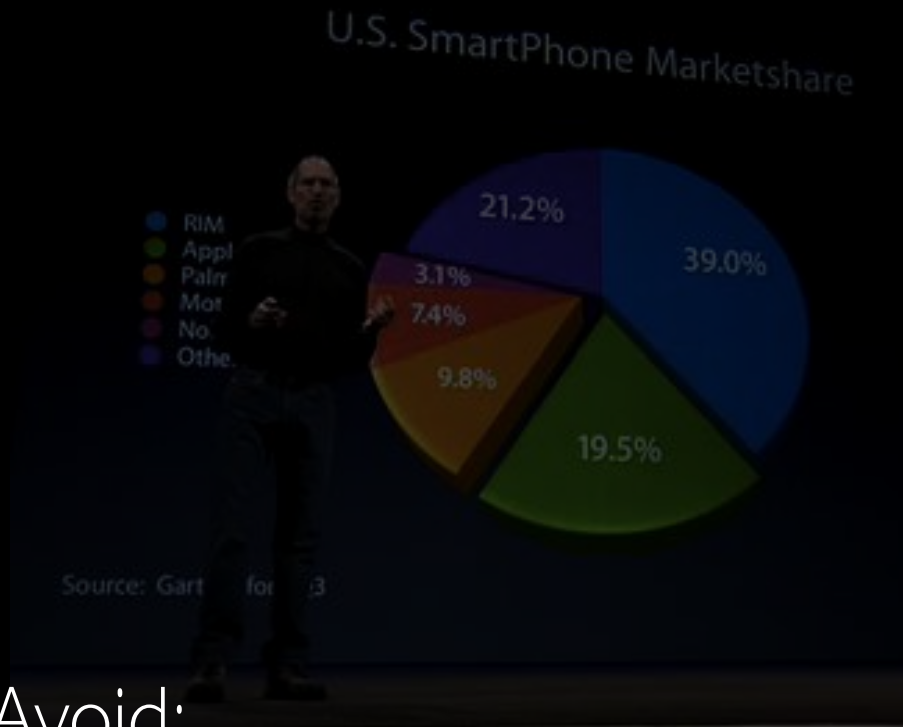
WHERE to BEGIN?



Data Visualization Considerations

- Chart type
- Data context
- Proper scales
- Avoid clutter
- Color use

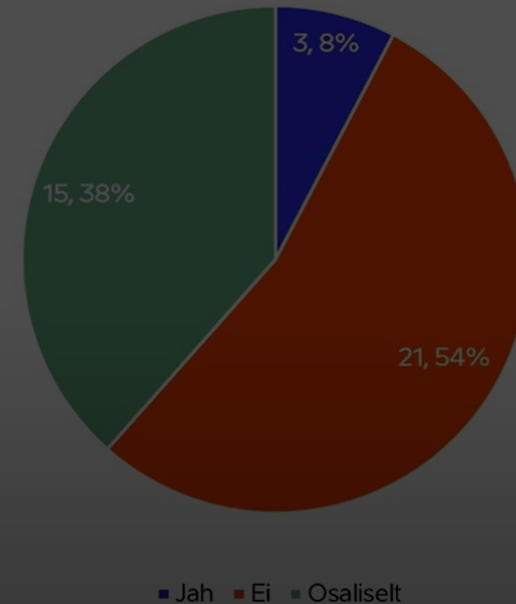
Chart Type



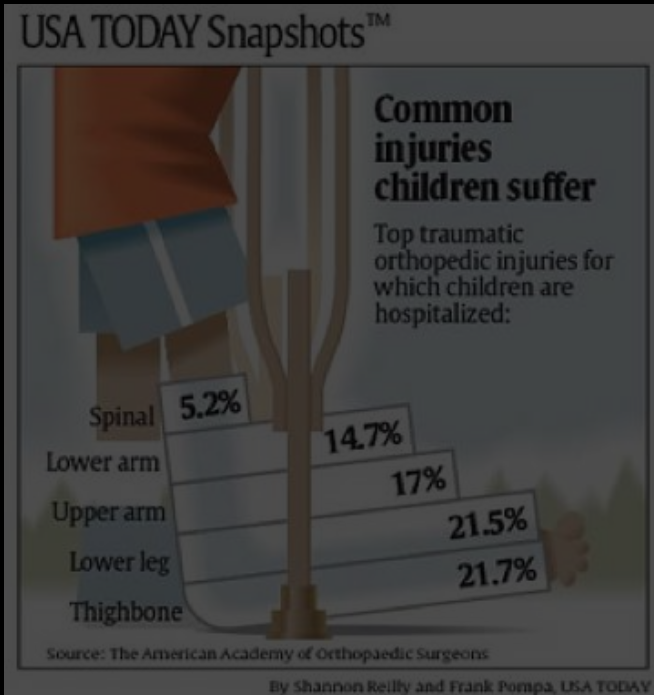
Avoid:

- 3D charts
- Pie charts

Kas asutus on kõik juurdepääsupiiranguteta andmed teinud koos kirjeldustega kättesaadavaks avaandmete teabeväras (vastavalt AvTS §29 lg 6)?

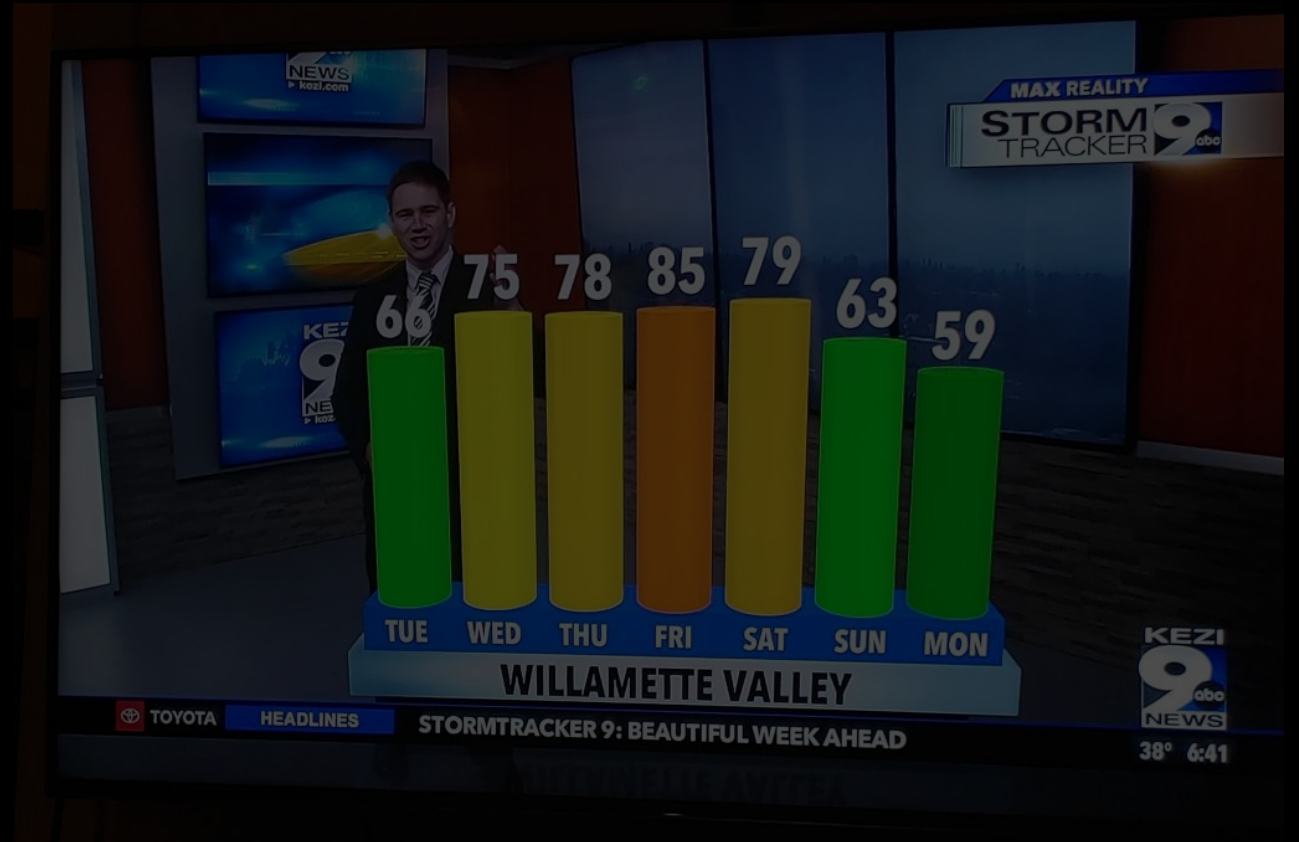


Data Context

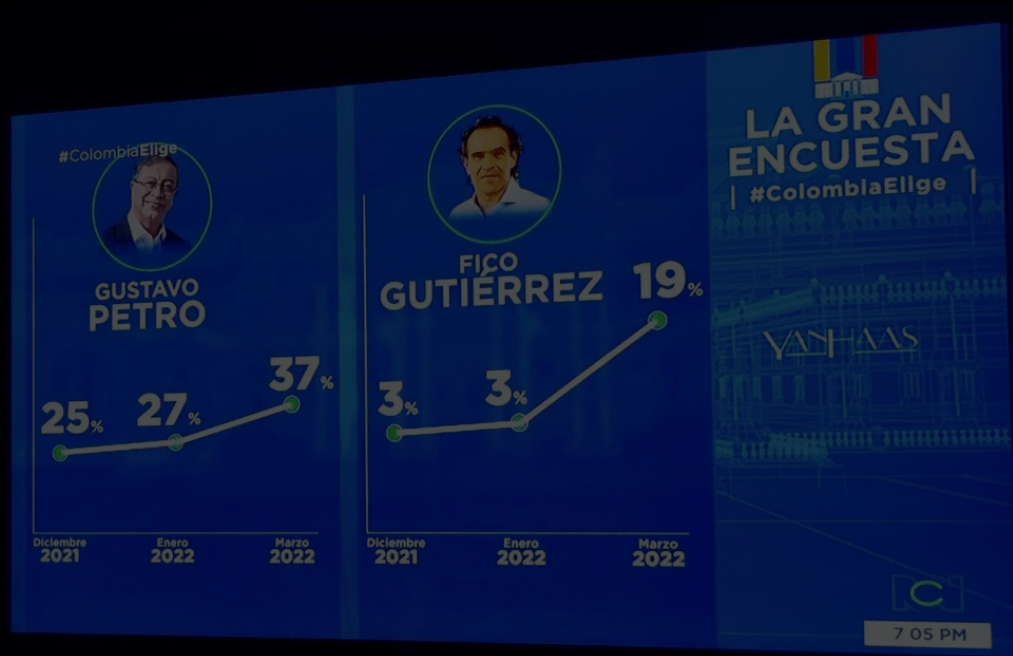


Add:

- Dataset scope
- Relevant labels

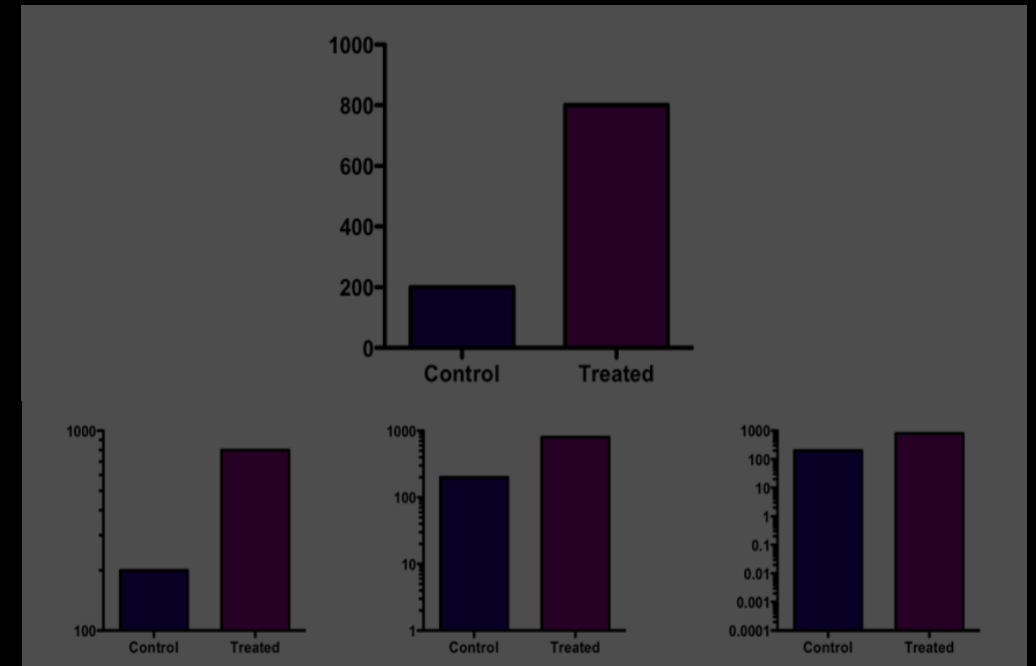


Proper Scales



Use:

- Zero baseline
- Linear scaling



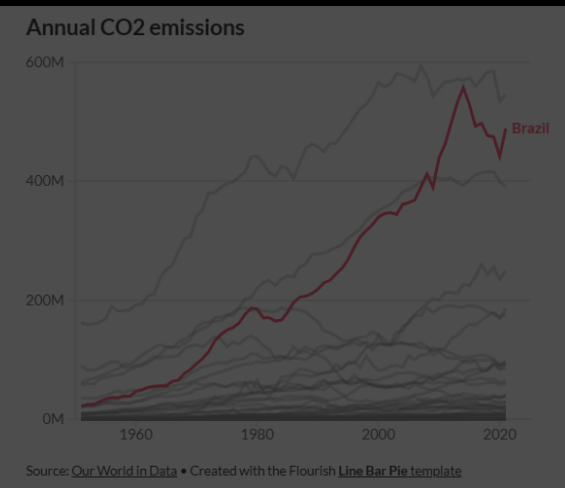
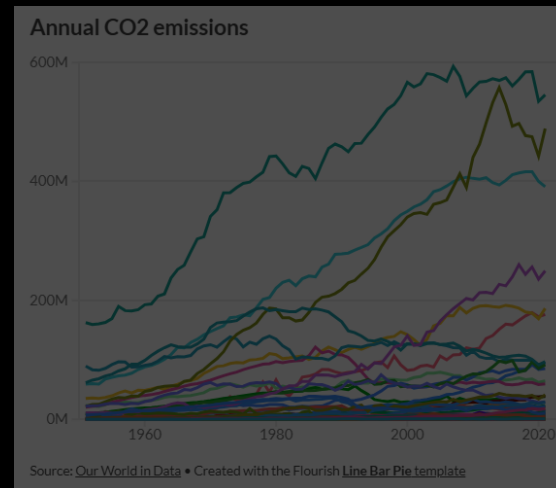
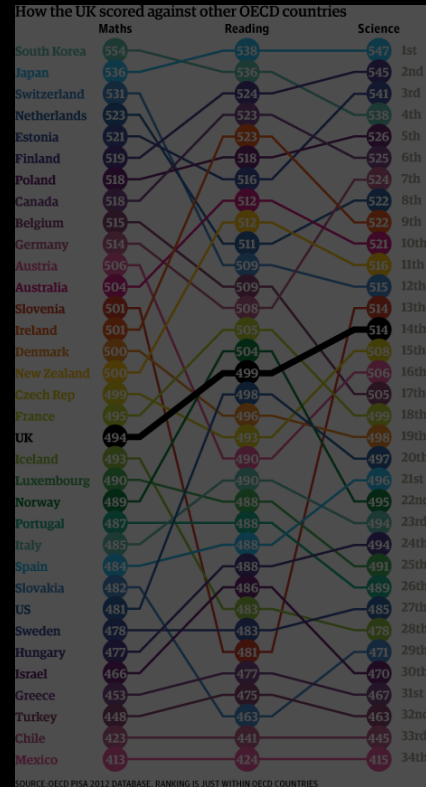
Clutter



Focus:

- One clear point
- Limit number of elements

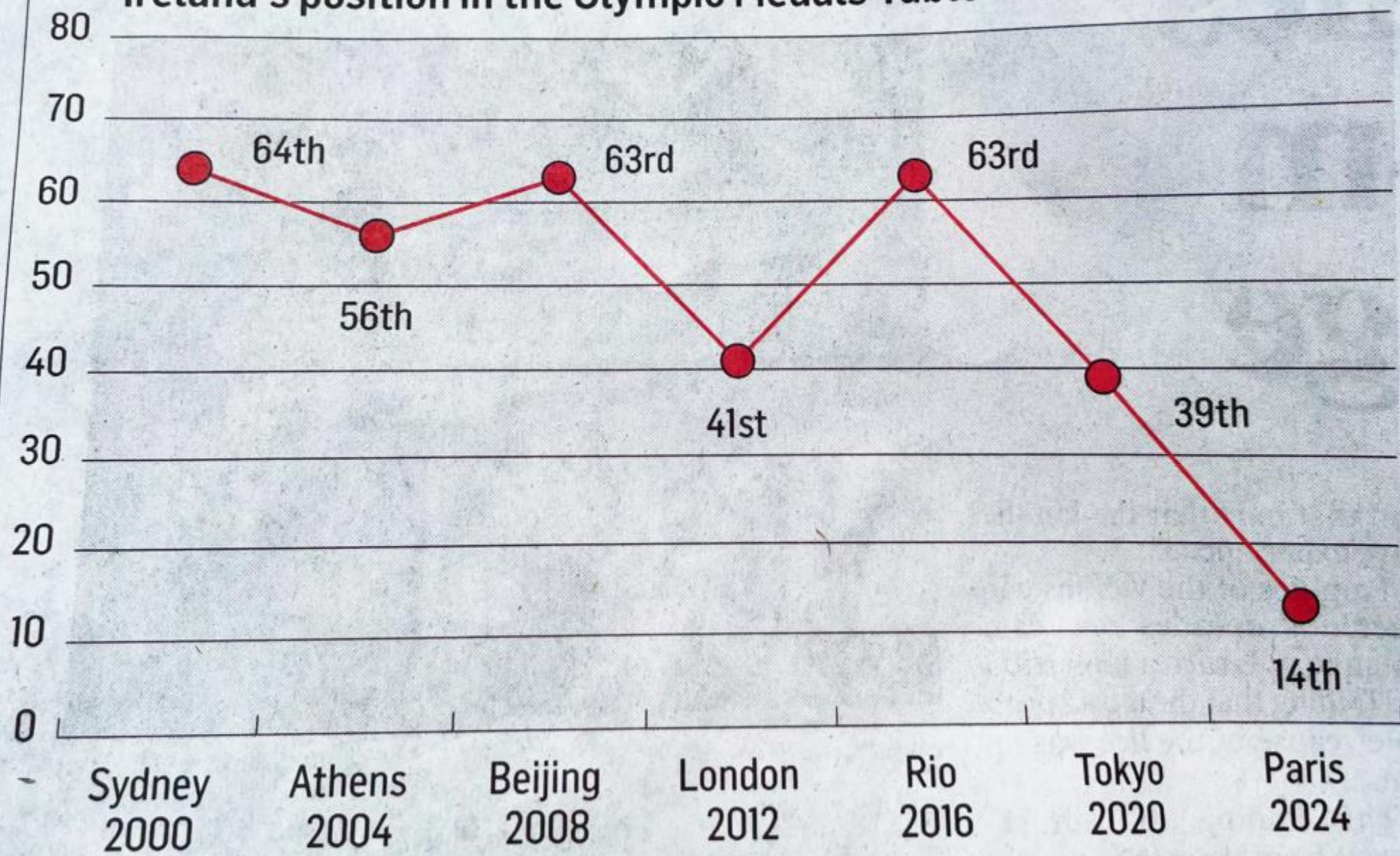
Color use



Use color:

- Highlighting
- Scale (when it makes sense)

Ireland's position in the Olympic Medals Table



Agenda

- ~~• Serving Data~~
- ~~• Data Visualization~~
- Quiz session

Further reading

- Chapter IX
 - Serving for ML
 - More ways to serve data
- Book

