

The main tools of data scientists:

- Measurement: Deals with how policies are constructed. I think that data scientists try to come up with measures or characteristics of some subjects or individuals that helps answer questions about policies. It's not always easy to come up with the right measure for policy or research questions. Thus, I think that data scientists need to have some creativity when it comes to choosing the right measure.
- Statistical programming languages: Python and R are the most common script languages in the data science field. I think though that R would be more useful for economists, but I think that one should also learn Python since it is widely used. I think though that prominent data scientists know how to work with enormous number of software and languages; so SPSS, SAS, Tableau,... are also important.

Visualization tools (usually included with the above): Packages in the mentioned script languages are available to construct different visuals. For instance, one can use *ggplot2* to create *geom_line* or *geom_point* figures.

- Big Data management software: use Resilient Distributed Datasets (Hadoop or Spark) to handle big data. SQL is also used to manage and transform data in the form of database.
- Data collection tools: Web scraping and API are mostly used to extract data from the net.
- Modeling: using data to make predictions or find causal relationships.