



INTRODUCTION TO TABLEAU

WHY TABLEAU?

- It can connect to a variety of data sources
 - Local Excel, csv, tab-delimited data files
 - Data servers
 - Statistical files including R, SAS, and SPSS data files
- It facilitates interactive data exploration and visualizations
- It requires very little programming knowledge (none if you are using basic functionalities)
- Visualizations can be “published” online and shared via blogs or other websites

VERSIONS

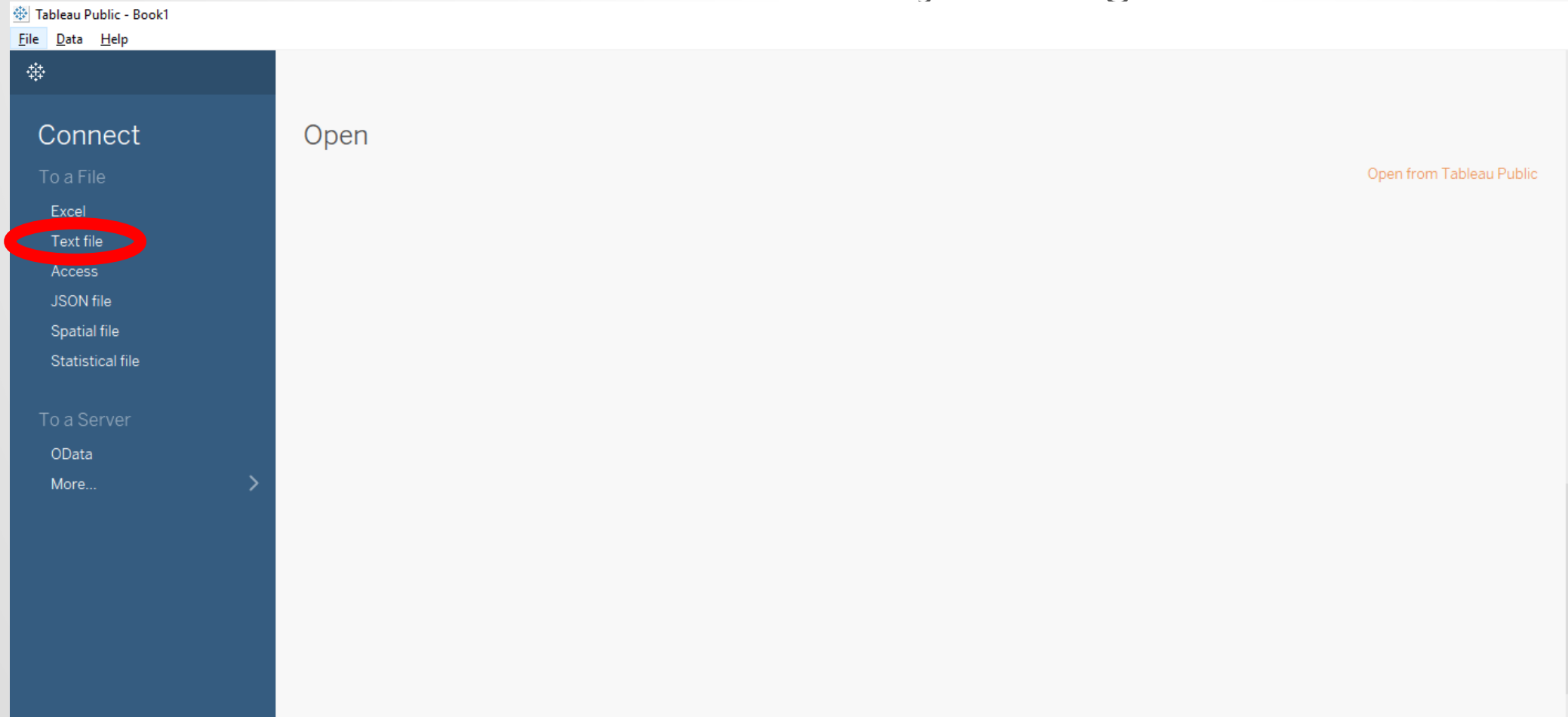
- Tableau Public
 - Free!
 - Workbooks are saved on Tableau Public's server rather than locally
- Tableau Desktop
 - Can save workbooks locally
 - More available data sources
 - Expensive in general, but free to faculty/students of accredited universities

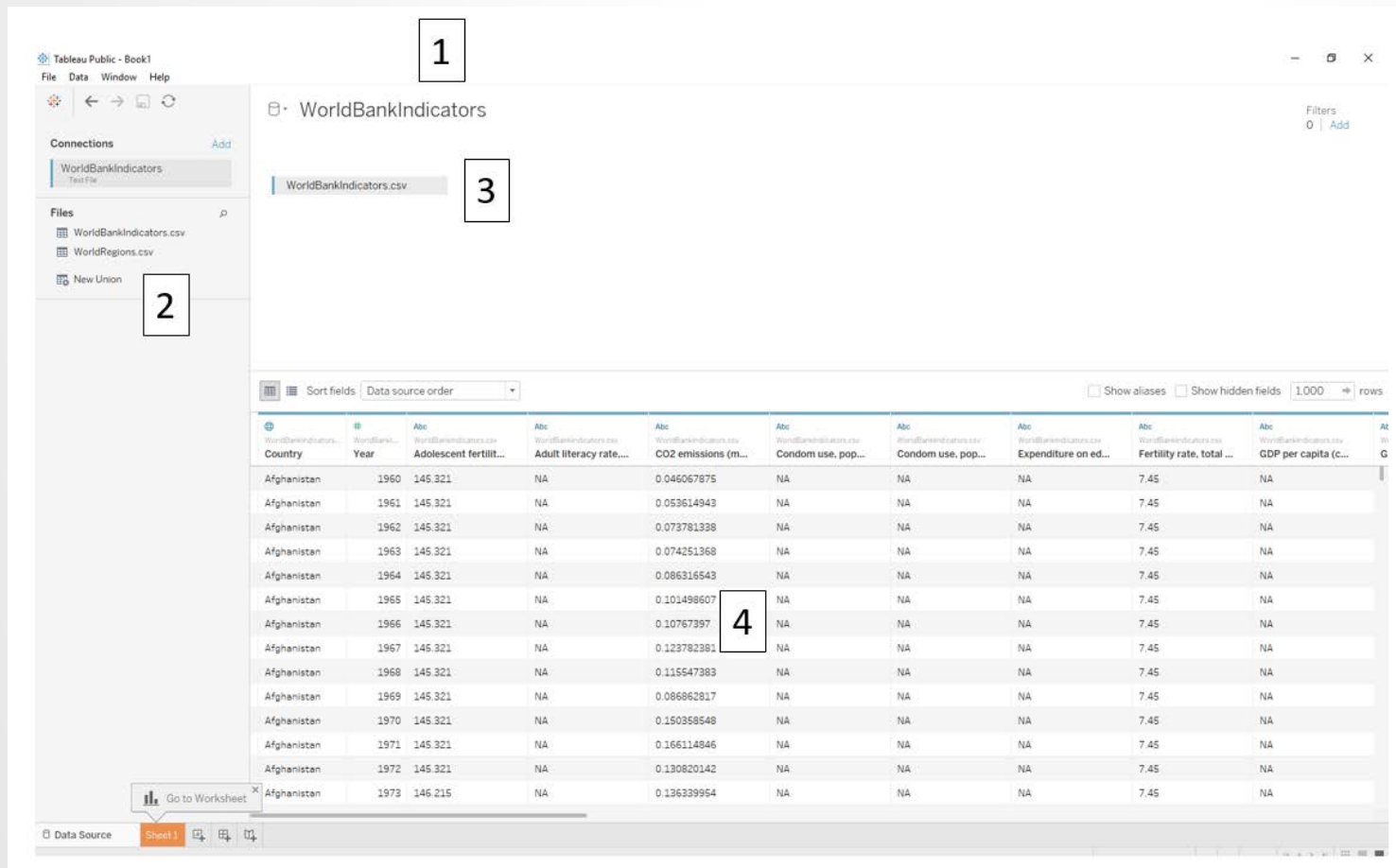
WHAT TABLEAU IS NOT....

- A data management tool! Essentially limited to...
 - New variable creation
 - Filtering
- R, Python, JMP are much better for heavy 'data wrangling'
 - Reshaping (wide to long to wide)
 - Complicated aggregation
- Do your heavy-duty cleaning outside Tableau; save to .csv or data file of choice

LET'S GET STARTED!

- The first time you open Tableau, you will see the "Home Page":
- Connect to the World_Bank.csv file, by clicking on the "Text File" option





1. The data file you are currently connected to. Click to rename it within Tableau.
2. Data sheets in the same working directory of the same file extension.
 - This is also where names of multiple sheets will show up, if you connect to an Excel file with multiple tabs.
3. The data sheets you want to visualize: you can click-and-drag sheets from "2" into this space to link multiple data sources.
4. A view of your data.

- Click on “Sheet 1”:

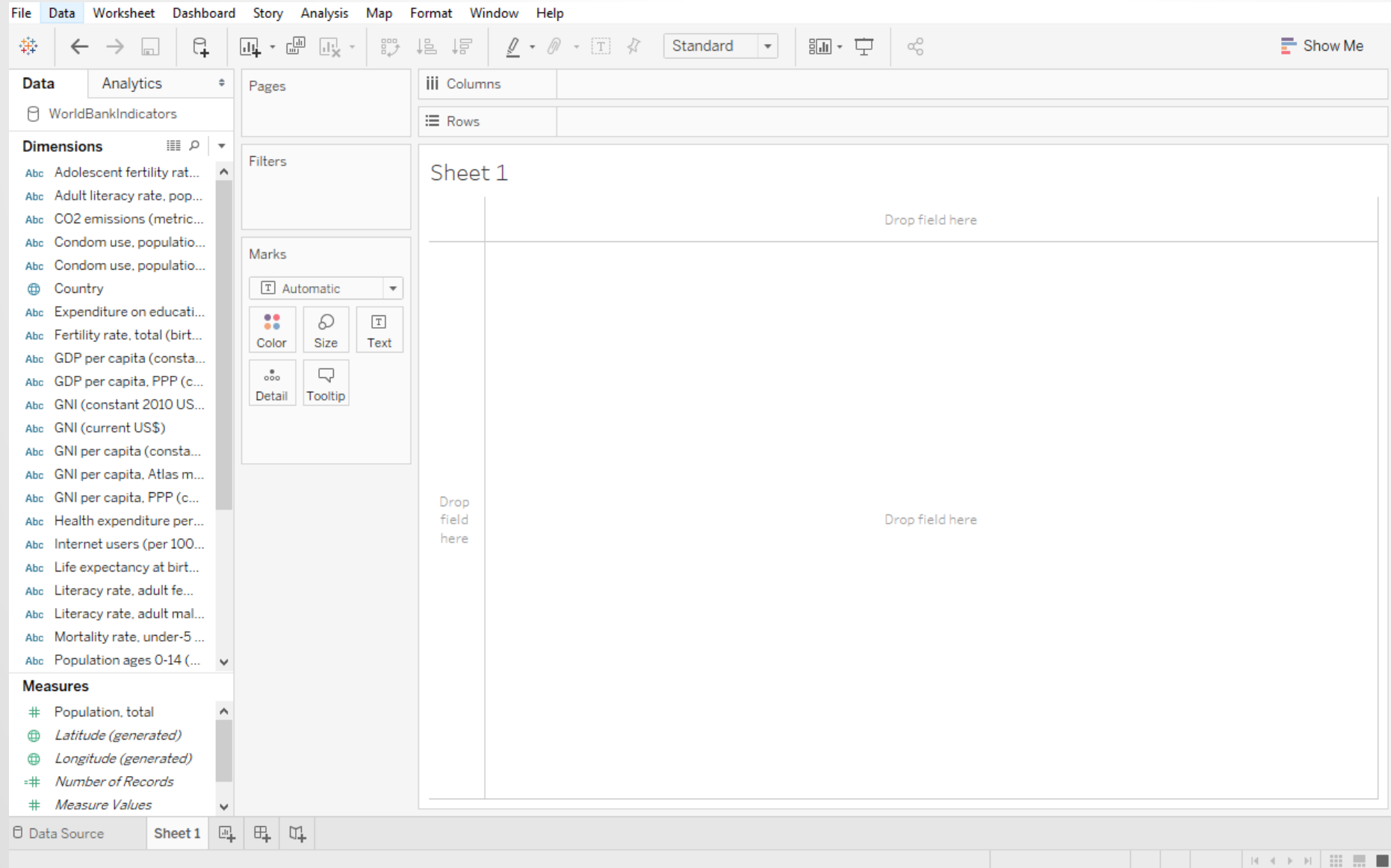
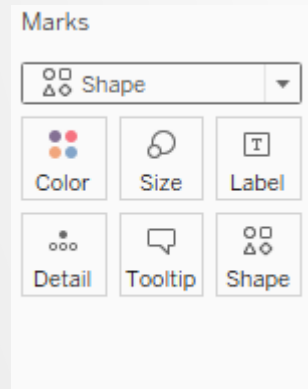


TABLEAU VOCABULARY

- Fields
 - I.e., “variables”; “data columns”
- Axes are determined by whether fields are **Discrete** or **Continuous**
 - Aka **Categorical** or **Quantitative**
 - **Discrete** fields create headers, or *discrete* bins. Blue pills.
 - **Continuous** fields create *continuous* AXES. Green pills
- Aggregation determined by whether fields are MEASURES or DIMENSIONS
 - Dimensions by default are NOT AGGREGATED
 - Measures by default are AGGREGATED (summed, averaged, etc.)

THE MARKS CARD



"X"

"Y"

Columns	
Rows	

- Dropdown menu to change geometries
- Drag-and-drop variables to map aesthetic attributes

- Note some fields not in the original data set:
 - *Measure Names*: Contains all the field names of the measures.
 - Vector of strings, containing all the names of the measures
 - In this case, *Measure Names* contains the strings
 - *Number of Records*: Essentially a column of 1's
 - Very useful for counting rows
 - *Measure Values*: All the numeric values of all measures.
 - Super long vector, with all data values of measures in the data set
 - Useful for blending multiple fields in the same graph
 - Latitude, Longitude: automatically generated if there is a geographic field (Country, here)

- Tableau is usually good at guessing whether fields should be discrete or continuous
- All the 'quantitative' fields in the .csv file contained NAs, which were interpreted as strings
- Change these to decimals, then to continuous, by highlighting and right-clicking:

- Note how this changes the way the missing values are interpreted, and format of decimals

WorldBankIndicators.csv Country	WorldBankL... Year	WorldBankIndicators.csv Adolescent fertilit...	WorldBankIndicators.csv Adult literacy rate...	WorldBankIndicators.csv CO2 emissions (m...	WorldBankIndicators.csv Condom use, pop...	WorldBankIndicators.csv Condom use, pop...	WorldBankIndicators.csv Expenditure on ed...	WorldBankIndicators.csv Fertility rate, total ...	WorldBankIndicators.csv GDP per capita (c...
Afghanistan	1960	145.321	NA	0.046067875	NA	NA	NA	7.45	NA
Albania	1960	54.4408	NA	1.258194928	NA	NA	NA	6.489	NA
Algeria	1960	123.8892	NA	0.553763578	NA	NA	NA	7.524	2466.798296
American Samoa	1960	NA	NA	NA	NA	NA	NA	NA	NA
Andorra	1960	NA	NA	NA	NA	NA	NA	NA	NA
Angola	1960	234.684	NA	0.104357101	NA	NA	NA	7.379	NA
Antigua and Barbuda	1960	126.144	NA	0.67061685	NA	NA	NA	4.425	NA
Argentina	1960	61.378	NA	2.367473032	NA	NA	NA	3.109	5852.538357
Armenia	1960	42.8244	NA	NA	NA	NA	NA	4.55	NA
Aruba	1960	106.2062	NA	NA	NA	NA	NA	4.82	NA
Australia	1960	44.8286	NA	8.582936643	NA	NA	NA	3.453	19300.33858
Austria	1960	48.3914	NA	4.373318828	NA	NA	NA	2.69	13087.94338
Azerbaijan	1960	43.221	NA	NA	NA	NA	NA	5.571	NA
Bahamas, The	1960	91.8424	NA	3.74983109	NA	NA	NA	4.495	13895.26086

Fields as discrete strings →

WorldBankIndicators.csv Country	WorldBankL... Year	WorldBankIndicators.csv Adolescent fertilit...	WorldBankIndicators.csv Adult literacy rate...	WorldBankIndicators.csv CO2 emissions (m...	WorldBankIndicators.csv Condom use, pop...	WorldBankIndicators.csv Condom use, pop...	WorldBankIndicators.csv Expenditure on ed...	WorldBankIndicators.csv Fertility rate, total ...	WorldBankIndicators.csv GDP per capita (c...
Afghanistan	1960	145.321	null	0.0461	null	null	null	7.45000	null
Afghanistan	1961	145.321	null	0.0536	null	null	null	7.45000	null
Afghanistan	1962	145.321	null	0.0738	null	null	null	7.45000	null
Afghanistan	1963	145.321	null	0.0743	null	null	null	7.45000	null
Afghanistan	1964	145.321	null	0.0863	null	null	null	7.45000	null
Afghanistan	1965	145.321	null	0.1015	null	null	null	7.45000	null
Afghanistan	1966	145.321	null	0.1077	null	null	null	7.45000	null
Afghanistan	1967	145.321	null	0.1238	null	null	null	7.45000	null
Afghanistan	1968	145.321	null	0.1155	null	null	null	7.45000	null
Afghanistan	1969	145.321	null	0.0869	null	null	null	7.45000	null
Afghanistan	1970	145.321	null	0.1504	null	null	null	7.45000	null
Afghanistan	1971	145.321	null	0.1661	null	null	null	7.45000	null
Afghanistan	1972	145.321	null	0.1308	null	null	null	7.45000	null
Afghanistan	1973	146.215	null	0.1363	null	null	null	7.45000	null

Fields as continuous decimals
(NAs appropriately recognized as such)

