

Wrangling Your Data

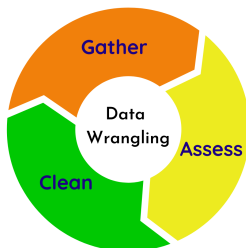
Kevin Donovan

UNC-Chapel Hill and IBIS Network

11/12/2020

Data in given form often not suitable for analysis

- Create new variables
- Rearrange variable order
- Rearrange row order
- Remove observations



Benefits of using R

- Can edit temporary versions of data
- All in code; transparent/reproducible
- Easily save output in various formats
- Very powerful tools

Wrangling Your Data

R Data Structure

Standard Format: **Data Frame**

country	year	cases	population
Afghanistan	1999	745	19984071
Afghanistan	2000	866	20095360
Brazil	1999	3737	17206362
Brazil	2000	8488	17404898
China	1999	21258	127291272
China	2000	21676	128042583

variables

country	year	cases	population
Afghanistan	1999	745	19984071
Afghanistan	2000	866	20095360
Brazil	1999	3737	17206362
Brazil	2000	8488	17404898
China	1999	21258	127291272
China	2000	21676	128042583

observations

country	year	cases	population
Afghanistan	99	745	19984071
Afghanistan	00	866	20095360
Brazil	99	3737	17206362
Brazil	00	8488	17404898
China	99	21258	127291272
China	00	21676	128042583

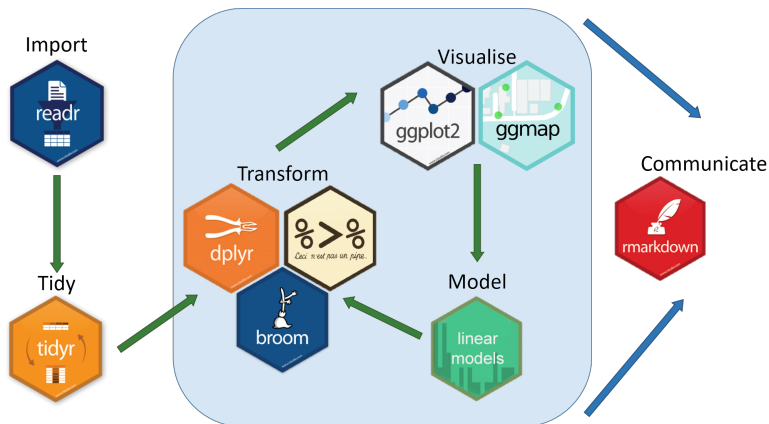
values

Data in **Tidy** form

Wrangling Your Data

Methods in R

- Base R (built-in functions)
- Tidyverse packages



Song of the Session