Data Science Basics in R

Day 3: Exploratory data analysis

Goals for today

- Define descriptive statistics & exploratory data analysis
- Create your first data visualization in R
- Identify options for visualization in R, including ggplot2
- Get creative and have fun exploring datasets

Descriptive statistics

Goals for today

- Define descriptive statistics & exploratory data analysis
- Create your first data visualization in R
- Identify options for visualization in R, including ggplot2
- Get creative and have fun exploring datasets

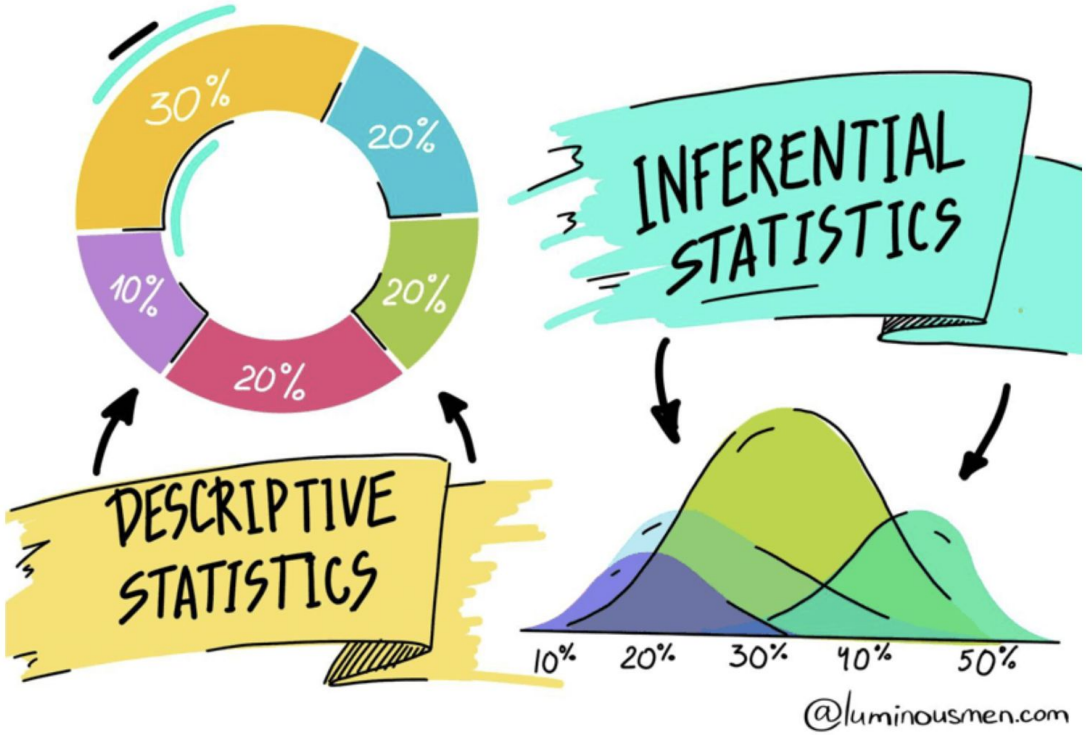
Descriptive statistics

Descriptive statistics summarize data, and typically describe three types of things:

- center (e.g., mean, median)
- spread (e.g., standard deviation, interquartile range)
- counts & rates (e.g., summary tables)

In a typical data analysis workflow, we explore these first! It's helpful to better understand your data, and to identify potential surprises.

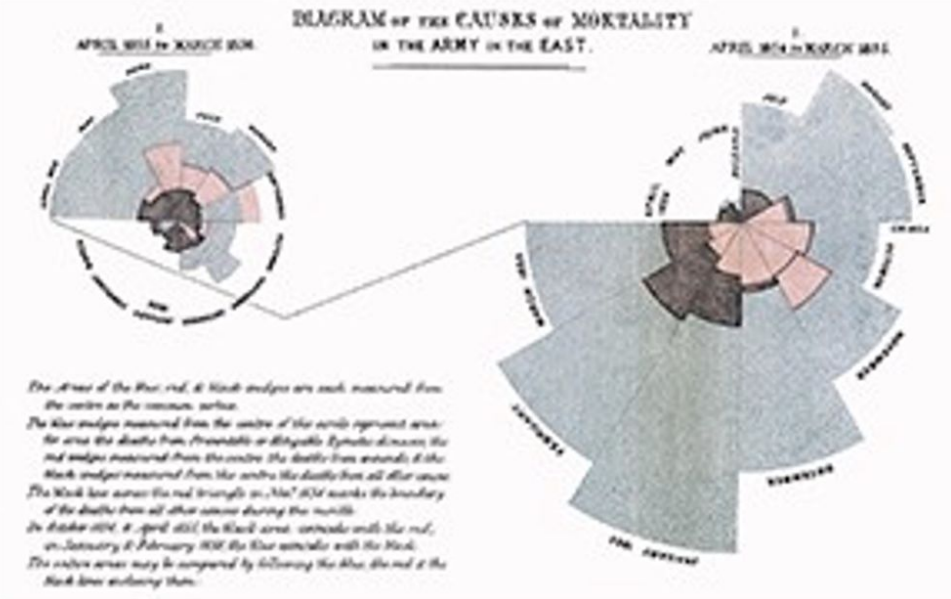
Descriptive statistics



Graphic by luminousmen.com
https://luminousmen.com/post/descriptive-and-inferential-statistics

Goals of data visualization

Florence Nightingale's polar area chart, “Diagram of the causes of mortality in the army in the East”.



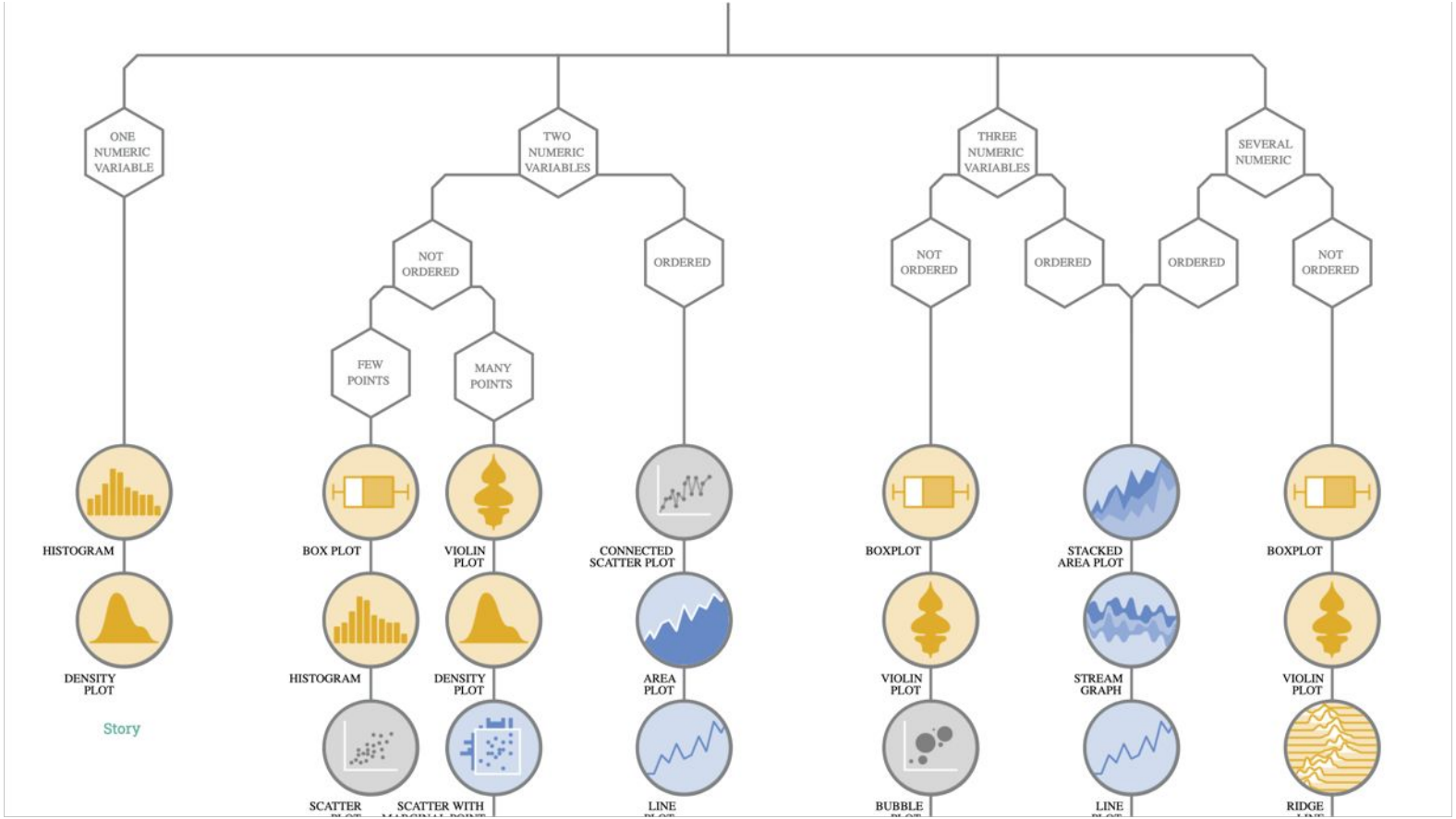
Significance, Volume 17, Issue 2, April 2020, Pages 26–30,
https://doi.org/10.1111/1740-9713.01376

Goals of data visualization

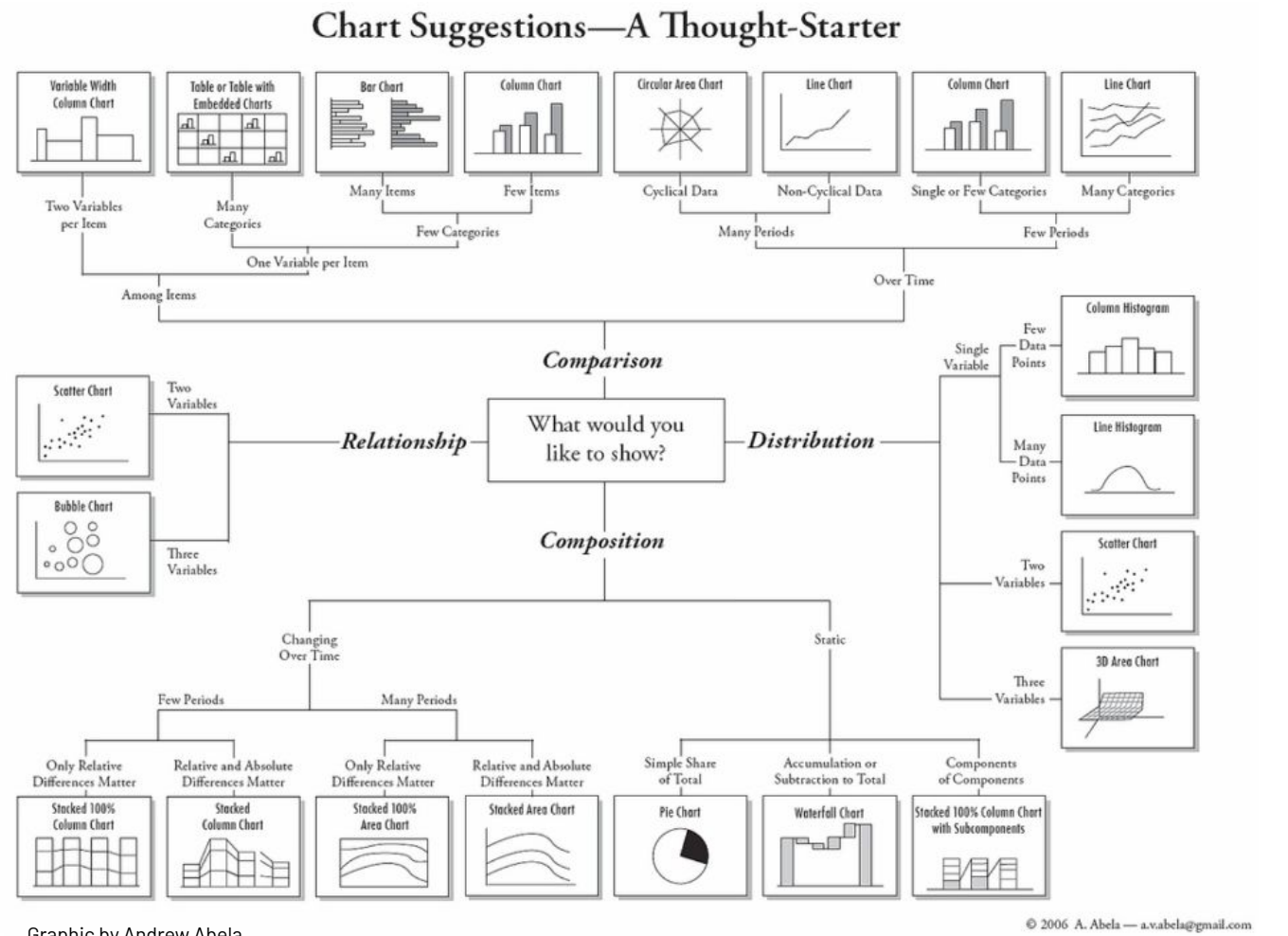
What makes a data visualization *good*?

Choosing a data visualization

(we'll talk more about this tomorrow)



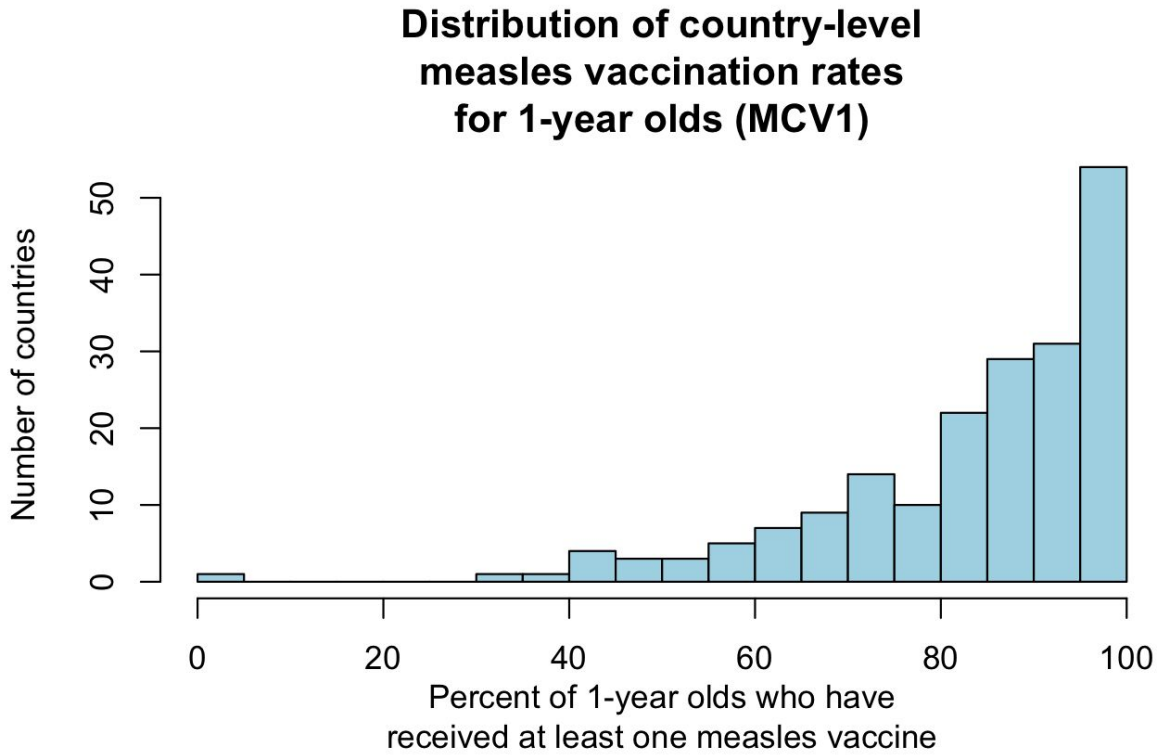
Graphic by Data to Viz
https://www.data-to-viz.com/



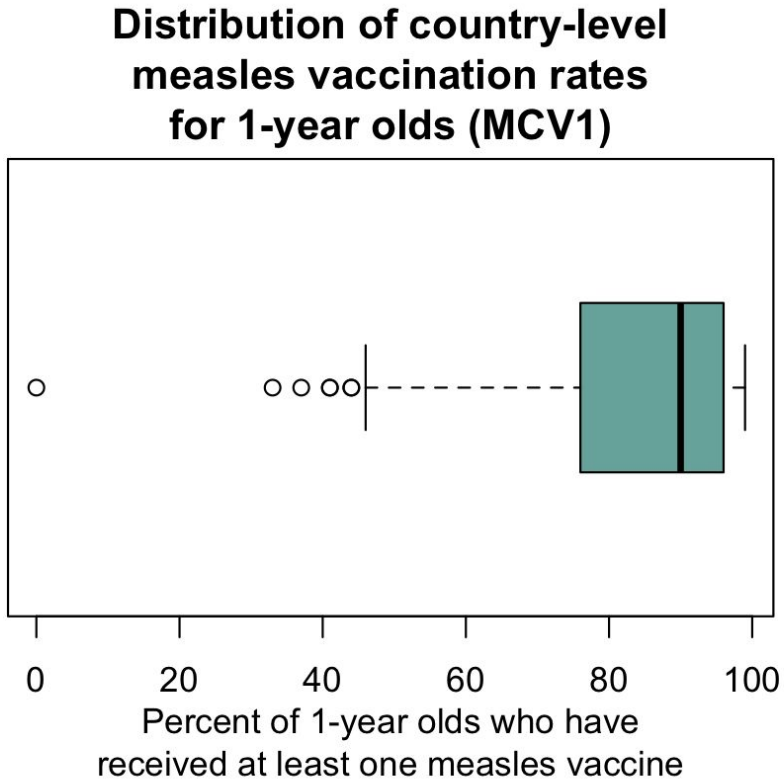
Graphic by Andrew Abela
https://extremepresentation.typepad.com/blog/2006/09/choosing_a_good.html

Plots in base R
(code in github for live demo)

Histogram

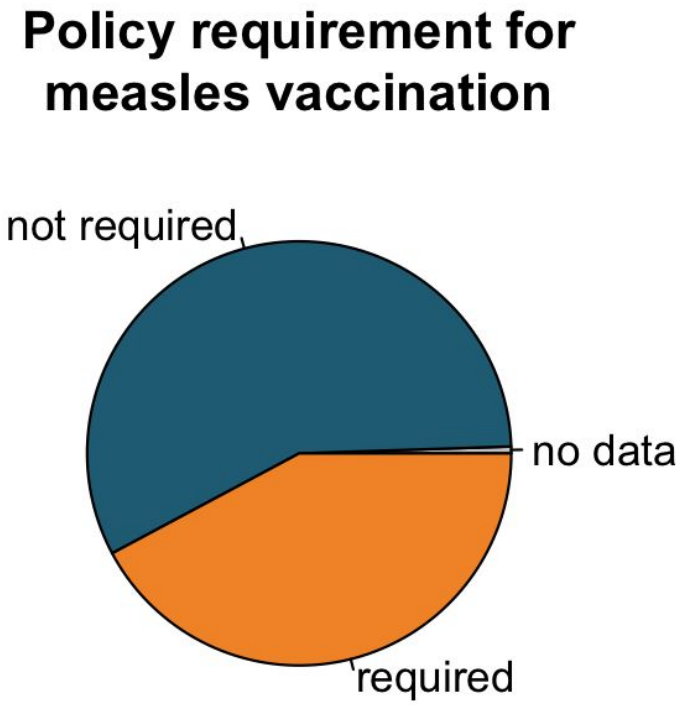

Distribution of country-level
measles vaccination rates
for 1-year olds (MCV1)
Number of countries
0
10
20
30
40
50
0
20
40
60
80
100
Percent of 1-year olds who have
received at least one measles vaccine

Boxplot

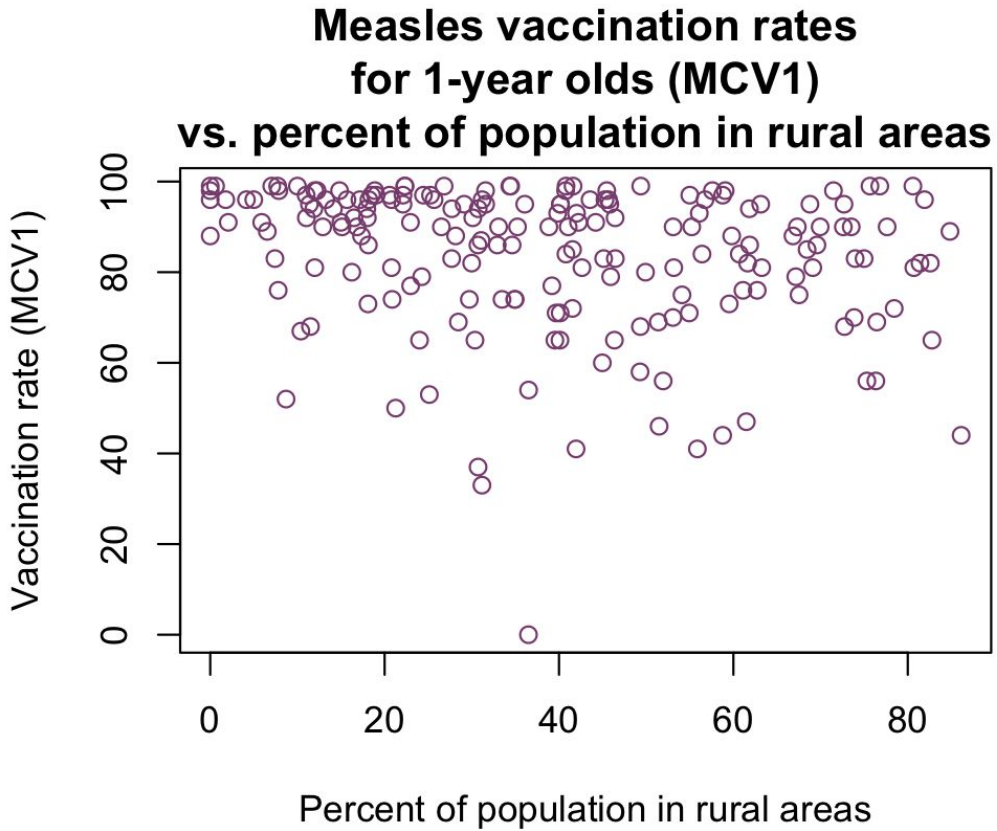

Distribution of country-level
measles vaccination rates
for 1-year olds (MCV1)
0
20
40
60
80
100
Percent of 1-year olds who have
received at least one measles vaccine

Barchart

Pie chart


Policy requirement for measles vaccination
not required
no data
required

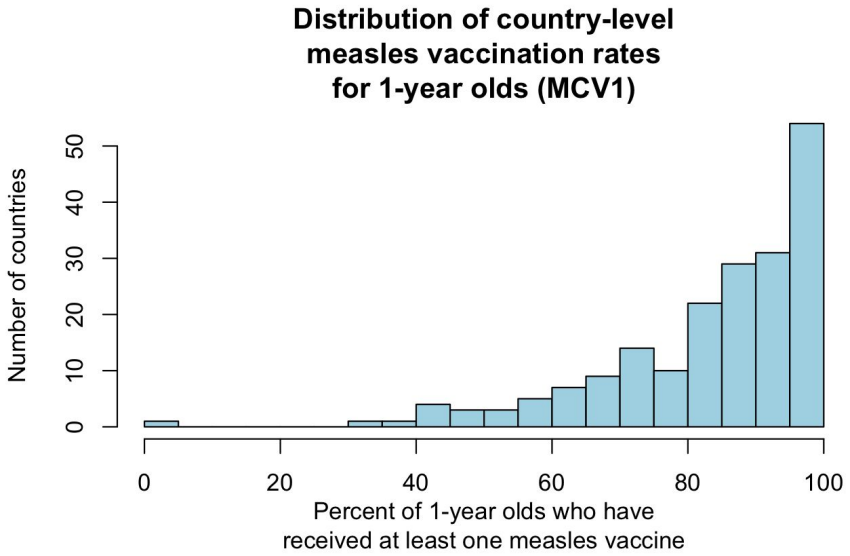
Scatterplot



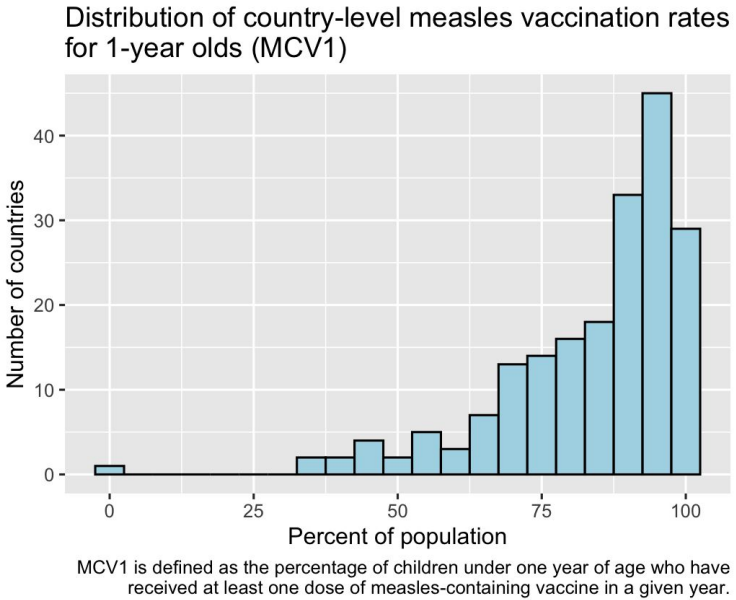
10 minute break

Plots in ggplot2
(code in github for live demo)

Histogram

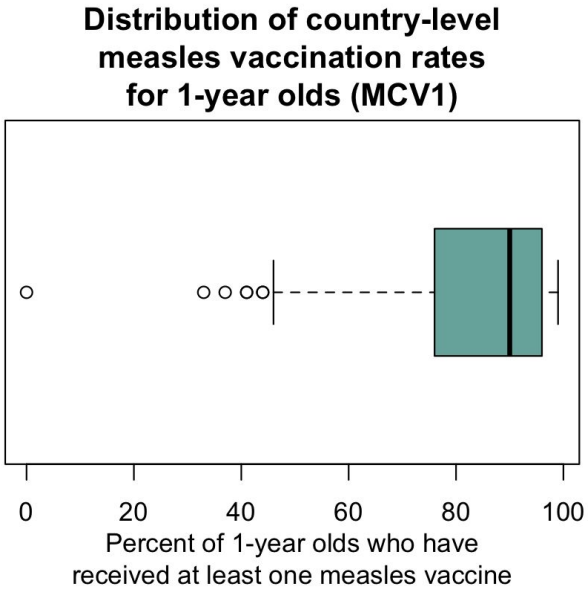


Base R

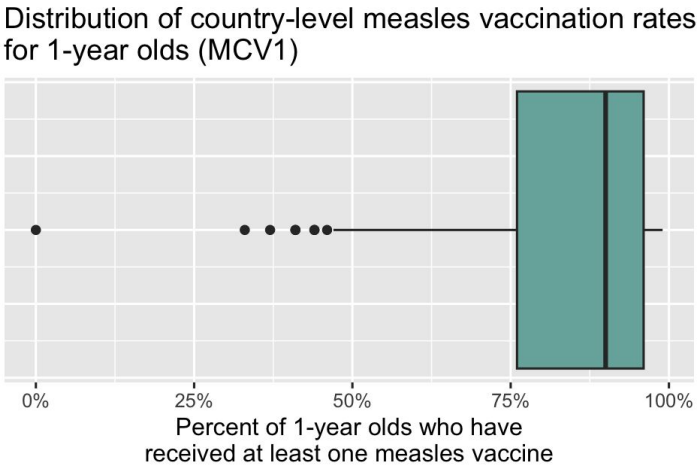


ggplot

Boxplot



Base R



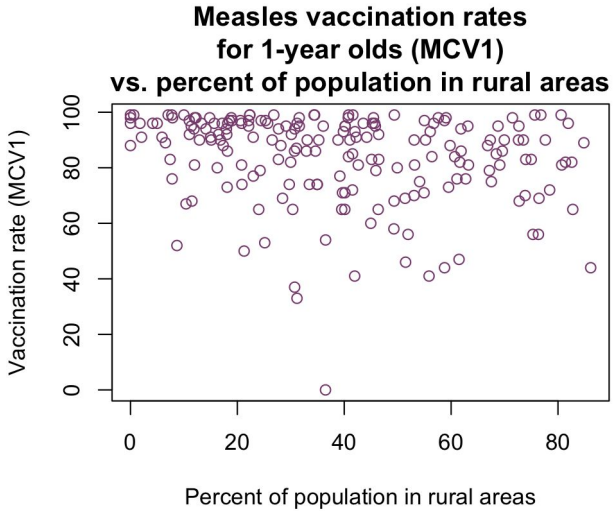
ggplot

Bar chart

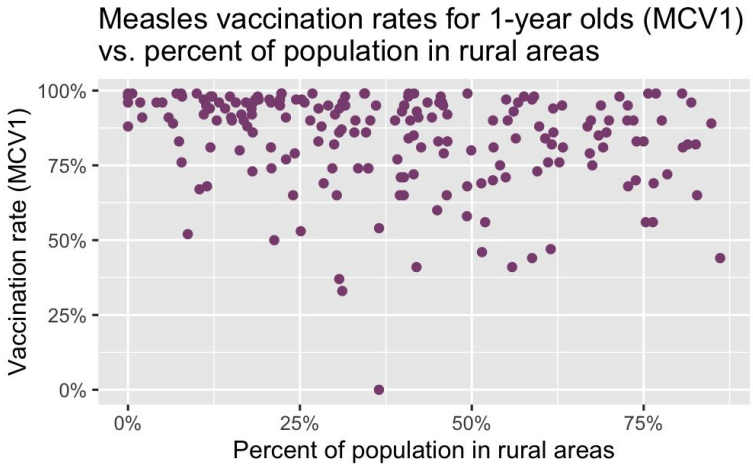
Base R

ggplot

Scatterplot



Base R



ggplot

Thank you!

See you tomorrow.

Please come with a fully charged laptop.