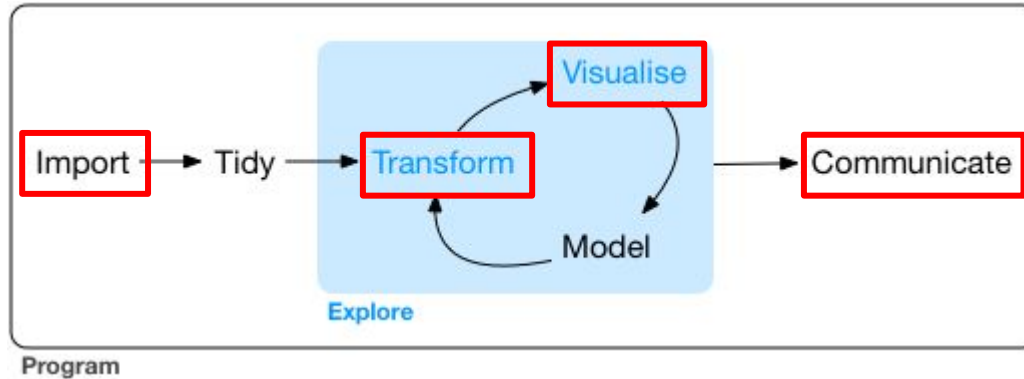




EXPLORATORY DATA ANALYSIS (EDA)

STEPS IN DATA ANALYSIS



Source: Wickham, Hadley, and Garrett Grolemund. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. First edition, O'Reilly, 2016. URL: <https://r4ds.hadley.nz/exploratory-data-analysis.html>

WHAT IS EDA?

EDA is an iterative cycle. You:

1. **Generate questions** about your data.
2. Search for answers by **visualising**, **transforming**, and **modelling** your data.
3. Use what you learn to refine your questions and/or generate new questions.

 **EDA is a basic tool for all other types of analysis!**

Source: Wickham, Hadley, and Garrett Grolemund. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. First edition, O'Reilly, 2016. URL: <https://r4ds.hadley.nz/exploratory-data-analysis.html>

TYPICAL TASKS IN EDA

- Become familiar with the data
- Assess data quality and fit for purpose
- Iteratively **transform** (**dplyr**) and **visualize** (**ggplot**) the data to find interesting **signals**
- Formulate **hypotheses** based on these signals
- **Collect evidence** to decide upon further investigation to confirm or refute hypothesis (inferential)

Source: Wickham, Hadley, and Garrett Grolemund. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. First edition, O'Reilly, 2016. URL:

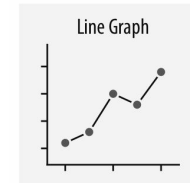
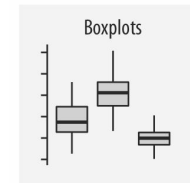
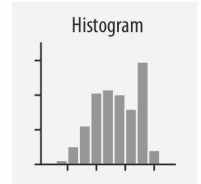
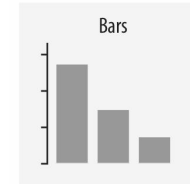
<https://r4ds.hadley.nz/exploratory-data-analysis.html>

GENERATE QUESTIONS

All questions are allowed. The **more the better**. Consider these as starting points:

- What type of **variation** occurs within my variables? E.g.
 - What is the distribution of values?
 - How does the variable develop over time?
- What type of **covariation** occurs between my variables?
 - What is the distribution of values by another (discrete) variable?
 - How do two variables change together? Are there any patterns?

Which is best for what?



<https://clauswilke.com/dataviz/directory-of-visualizations.html>

EXAMPLE: VIDEO GAME STUDY

EDA IN PRACTICE

- Let's look at a real-world scientific study as an example for EDA:
 - [Website with all code](#) (made with RMarkdown / **bookdown**)
 - [Preprint of paper](#)
 - [Supplemental material](#)
 - Nintendo / Survey data in [R-Project on GitHub](#)

