

Lesson 3:  
Assessing Data

SEARCH

RESOURCES

CONCEPTS

✓ 2. Lesson Outline

3. Dataset: Oral Insulin Phase II Cli...

4. Unclean Data: Dirty vs. Messy 1

5. Unclean Data: Dirty vs. Messy 2

6. Assessment: Types vs. Steps

7. Visual Assessment

8. Visual Assessment: Acquaint Yo...


9. Quality: Visual Assessment 1


10. Assessing vs. Exploring

11. Quality: Visual Assessment 2

12. Data Quality Dimensions 1

13. Data Quality Dimensions 2

 Mentor Help  
Ask a mentor on our Q&A platform

 Peer Chat 3  
Chat with peers and alumni



## Lesson Outline

Data wrangling process:

- Gather
- Assess (this lesson)
- Clean

**Assessing** your data is the second step in data wrangling. When assessing, you work, inspecting your dataset for two things: data quality issues (*i.e. content issues*) (*i.e. structural issues*).

Assessing is the precursor to cleaning. You can't clean something that you don't assess. In this lesson, you'll learn to identify and categorize common data quality and tidiness issues. This is the shortest and most "hands-off" code-wise of all four in the course because cleaning is relative to gathering and cleaning. We have tried to include quizzes to help you practice.

### This lesson will be structured as follows:

- You'll get motivated to assess (and later clean) the dataset for lessons 3 and 4. The dataset is a trial data that compares the efficacy and safety of a new oral insulin to treprostinil.
- You'll learn to distinguish between dirty data and messy data
- You'll assess the data visually and programmatically to identify:
  - Data quality issues
  - Tidiness issues
- You'll learn about data quality dimensions and categorize each of the data quality issues identified above into its appropriate dimension

To begin, I want to introduce you to the dataset you will be assessing in this lesson.