

Plan:

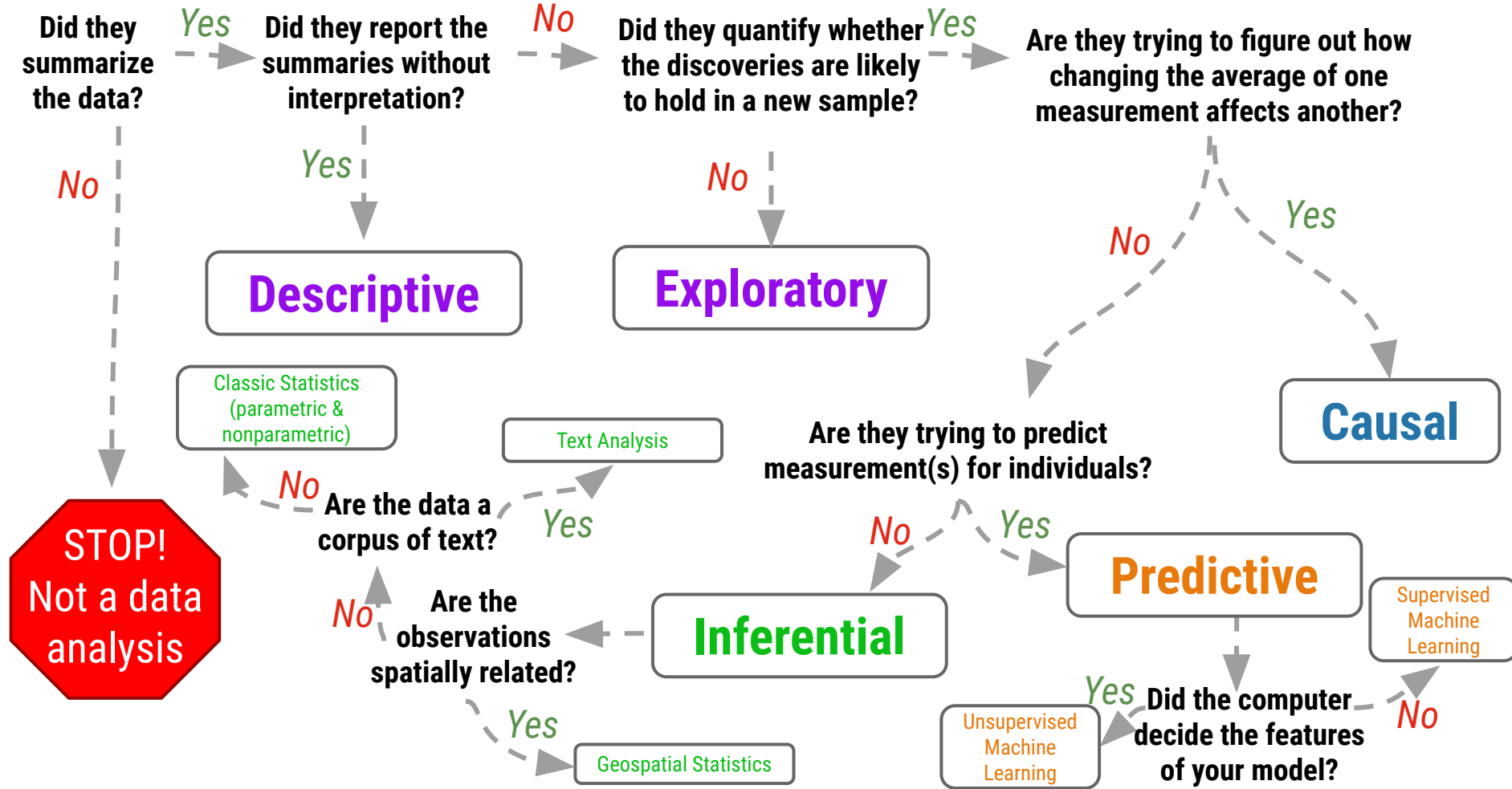
1. Introduce EDA
2. Explain goal of EDA

Exploratory Data Analysis

Shannon E. Ellis, Ph.D
UC San Diego



Department of Cognitive Science
sellis@ucsd.edu



Exploratory: The goal is to find unknown relationships between the variables you have measured in your data set. Exploratory analysis is open ended and designed to verify expected or find unexpected relationships between measurements.

Exploratory



Exploratory Data Analysis (EDA)
detective work answering the question:
“What can the data tell us?”

Why EDA?

- Understand data properties
- Discover Patterns
- Generate & Frame Hypothesis
- Suggest modeling strategies
- Check assumptions (sanity checks)
- Communicate results (present the data)

.....and if you don't, you'll regret it

Exploratory

The
dataset

You

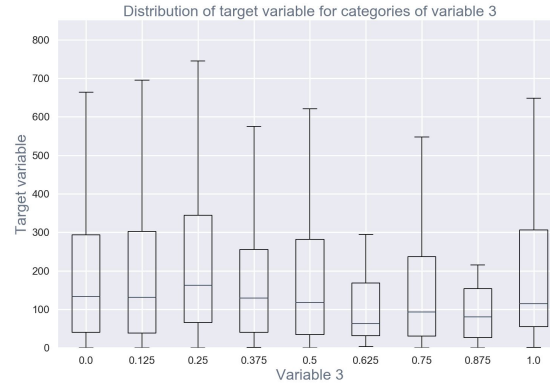
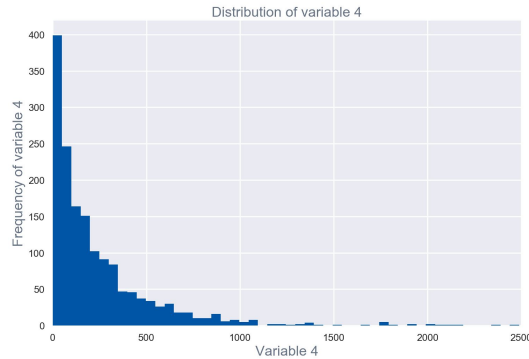


The general principles of exploratory analysis:

- Look for missing values
- Look for outlier values
- Calculate numerical summaries
- Generate plots to explore relationships
- Use tables to explore relationships
- If necessary, transform variables

EDA Approaches to “Get a Feel for the Data”

Understanding the relationship between variables in your dataset



Univariate

understanding a single variable
i.e.: histogram, densityplot, barplot

Bivariate

understanding relationship between 2 variables
i.e.: boxplot, scatterplot, grouped barplot, boxplot

Dimensionality Reduction

projecting high-D data into a lower-D space
i.e.: PCA, ICA, Clustering