The Data Analysis Pipeline,
The Big Picture of elementary inference,
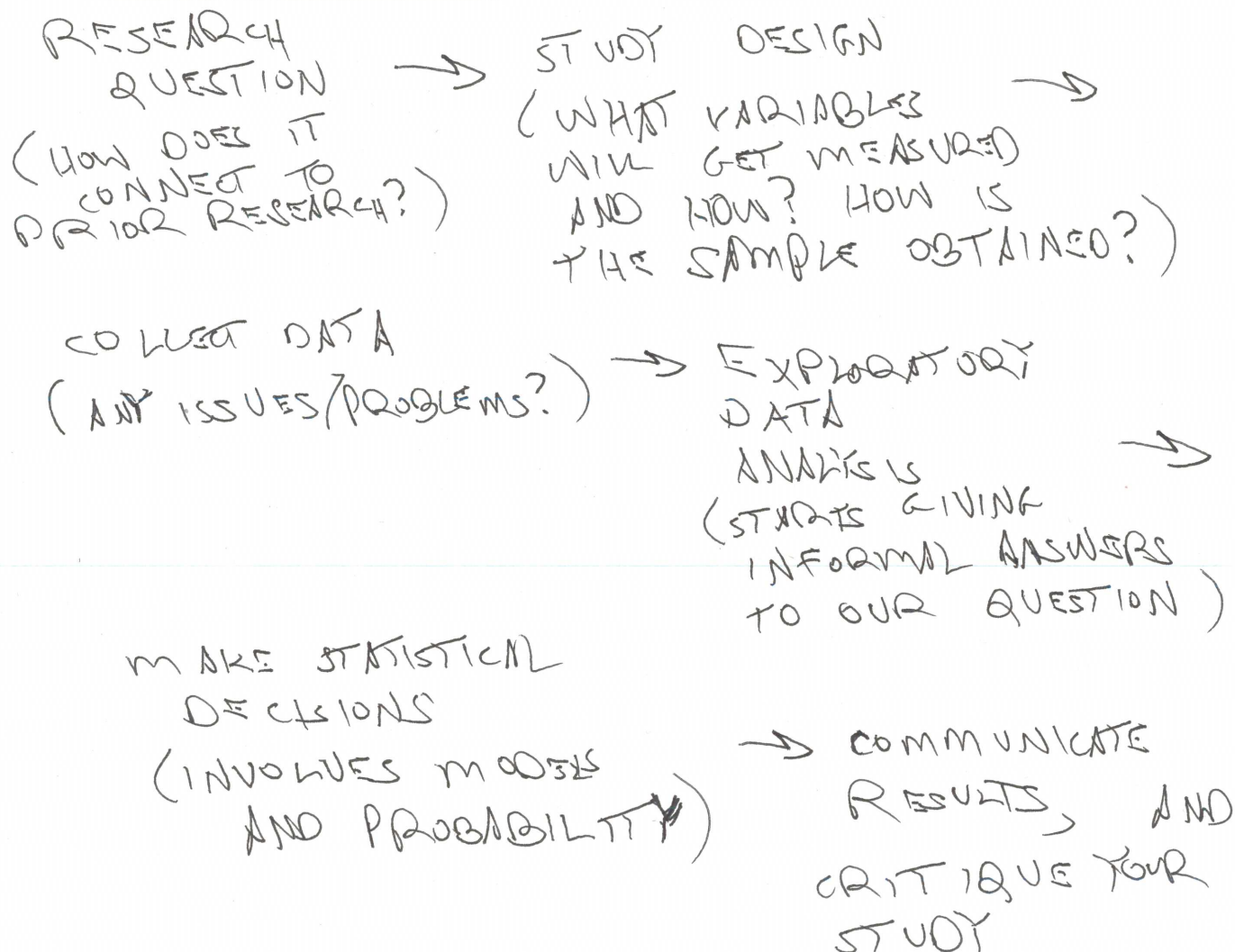# and
Identifying Data Types

Background

The word "statistics" derives from 'State,' originally referring to the data gathered to assist the administration of a Nation.

But today, the term more broadly encompasses the application of scientific principles for understanding and improving the world.

If research is the vehicle that takes us from ignorance to knowledge, then statistics is what's under the hood (and also how the car gets built). *Statistics is the technology of data.*
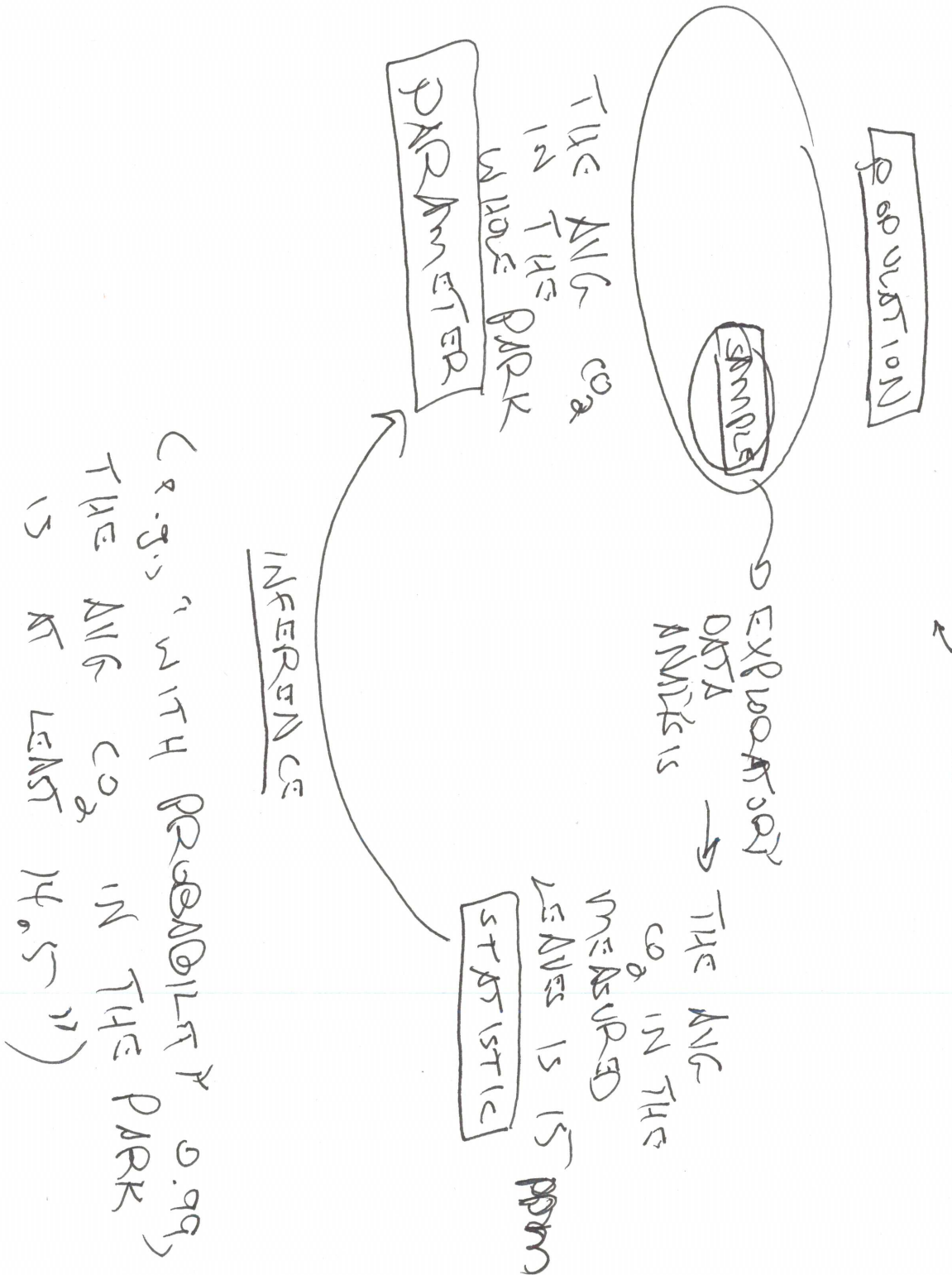
The Data Analysis 'pipeline':

RESEARCH QUESTION → STUDY DESIGN
(HOW DOES IT CONNECT TO PRIOR RESEARCH?)   (WHAT VARIABLES WILL GET MEASURED AND HOW? HOW IS THE SAMPLE OBTAINED?) →

COLLECT DATA (ANY ISSUES/PROBLEMS?) → EXPLORATORY DATA ANALYSIS (STARTS GIVING INFORMAL ANSWERS TO OUR QUESTION) →

MAKE STATISTICAL DECISIONS (INVOLVES MODELS AND PROBABILITY) → COMMUNICATE RESULTS, AND CRITIQUE YOUR STUDY

The 'Big Picture' of elementary inference

The 'Big Picture' of Inference: Example 1:
"What's the Carbon Dioxide (CO_2) level in the leaves on the trees in Schenley Park?"

POPULATION

SAMPLE

EXPLORATORY
DATA
ANALYSIS → THE AVG
CO_2 IN THE
MEASURED
LEAVES IS 15 ppm

STATISTIC

THE AVG CO_2
IN THE
WHOLE PARK

PARAMETER

INFERENCE

(e.g., "WITH PROBABILITY 0.99,
THE AVG CO_2 IN THE PARK
IS AT LEAST 14.5")