INTRODUCTION TO R AND R STUDIO

CLASS #2 | GEOG 215

Introduction to Spatial Data Science

Spring 2020

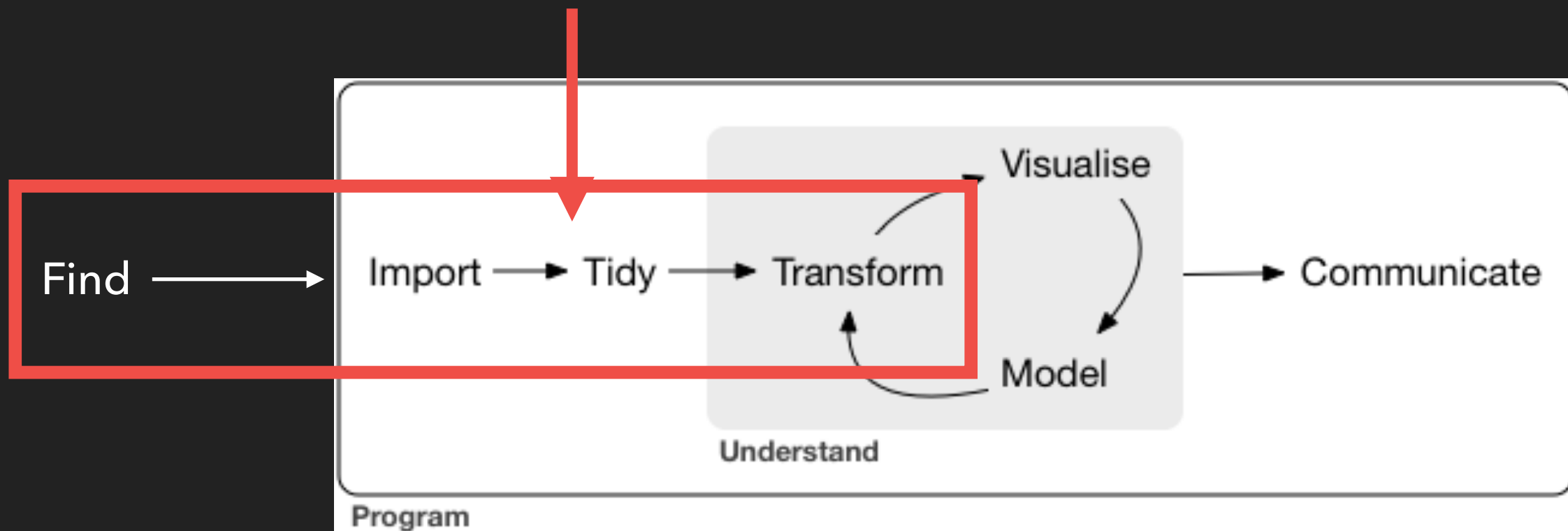# TODAY'S CLASS

▸ A Brief History of R (The what's and Why's)

▸ Rstudio Demonstration

▸ In class exercises (poll everywhere)

▸ Introduction to levels of measurements

▸ In-class practice with data types and data structures

▸ Glimpse at Lab 1

REMINDERS

▸ Register and complete the Poll everywhere survey

  ▸ https://pollev.com/goelvarun553/

▸ Sign up for class discussion on Piazza

  ▸ http://piazza.com/unc/spring2020/geog215/home

DATA SCIENCE PROCESS

"The most time consuming and least fun"

Find → Import → Tidy → Transform → Visualise → Model → Communicate

Understand

Program

Source: R for Data Science
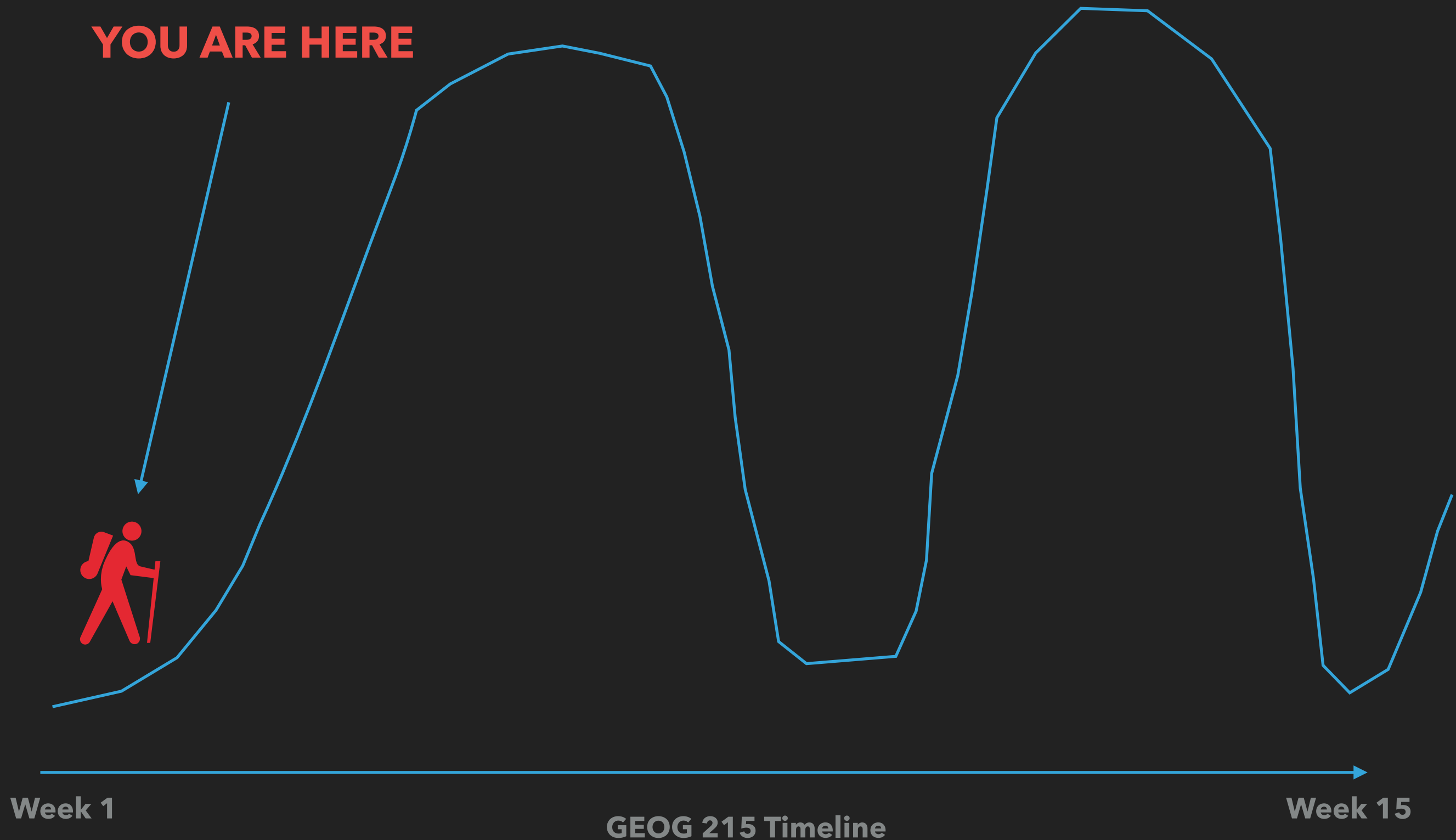
WHAT IS R?

▸ Software and a language

▸ S (1976) —-> R (1991) —> Public use (2000)

▸ Developed for Data Analysis

  ▸ Primarily for the 'user'

  ▸ To later become a 'developer'

WHY R

▸ Open Source

  ▸ Free as in Beer, but also Free as in Speech

▸ Year-round maintenance, frequent releases

▸ Publication quality graphics (first choice for academics)

▸ Vibrant user community (Over 20 years of community base)

▸ Powerful programming base to develop tools (Spatial data tools for example)

▸ Consistent grammar for Data wrangling (Next class)

WHY NOT R

▸ Memory Hog compared to other statistical packages

▸ Functionality based on consumer demand

▸ Slower compared to Python or Matlab

▸ Steep Learning curve (Some peculiarities compared to other languages)

NAVIGATING R STUDIO

▸ OPEN RStudio in your laptops

  ▸ We will go through the different panes of R studio

  ▸ Interact with the console

  ▸ Create our first R script

POLL EVERYWHERE

▸ OPEN your laptop and go to

  ▸ https://pollev.com/goelvarun553

  ▸ Respond to the 10 survey questions (individually)

OUR DATA COMES FROM SOMEWHERE, RIGHT ?

▸ Almost any measurement device can be used to measure attributes for Geographic Information

  ▸ eyes, remote sensor, weather station, flow gauge, survey of behavior or perceptions

▸ Every measurement or characteristic must also have units or scale associated with it

  ▸ e.g. , Distance = 10 ?

LEVELS OF MEASUREMENT

▸ **Nominal**

  ▸ Simplest type of attribute

  ▸ An identifier value - Name, letters, colors, numbers ?

  ▸ Question: what kind of mathematical operations can you apply to Nominal data ?

▸ **Ordinal**

  ▸ Values that have a natural order/rank

    ▸ What math operations can you use?

LEVELS OF MEASUREMENT

▸ **Interval**

   ▸ Numeric values (measurements) - e.g, temperature on Celsius scale

   ▸ Difference between values makes sense

   ▸ Arbitrary "0" measurement

▸ **Ratio**

   ▸ Numerical - ratios between values make sense

   ▸ E.g, population counts, % population with diseases, Air pollution rates

      ▸ Non-arbitrary "0" measurement

LEVELS OF MEASUREMENT

▸ Cyclic

   ▸ Numeric, but non- "0" based

   ▸ At the last value, cycles back to first

   ▸ E.g., months, latitude, longitude

PERMITTED MATHEMATICAL OPERATIONS

▸ Nominal - equivalence, count

▸ Ordinal - Greater than, less than

▸ Interval - difference, subtraction

▸ Ratio - multiplication, division

▸ Cyclical - special rules: e.g., compass directions

QUESTIONS SO FAR?

WORK THROUGH CODE EXAMPLES IN R

▸ Download file on website under lectures

   ▸ Basic data types

   ▸ Basic data structures

   ▸ Think about different levels of measurements and how they are stored

▸ Complete Swirl modules for week 2 if you haven't already

INTRO TO LAB 1

BEFORE NEXT CLASS

▸ Glance through/ start Lab 1 - will be uploaded tonight (Due Next Monday Jan 20 - 11:59 pm)

▸ Join Piazza (Participate)

▸ Practice, Practice, Practice

▸ Complete class survey about programming experience

QUESTIONS ?