Lecture 8:
Comparing Distributions

Announcements

1. Please return radiant thermometers to me now

Reminders:

1. HW#2 is due next class via eCampus

Let's open up R…

Differences

- Often we are concerned whether there are statistically significant difference between samples or between a sample and the larger population

- Difference if there is the means are different between samples or between the sample and the population

- Students t-test for small samples (<30) or z-test for large samples (>30)

Sample 1   **Sample 2**
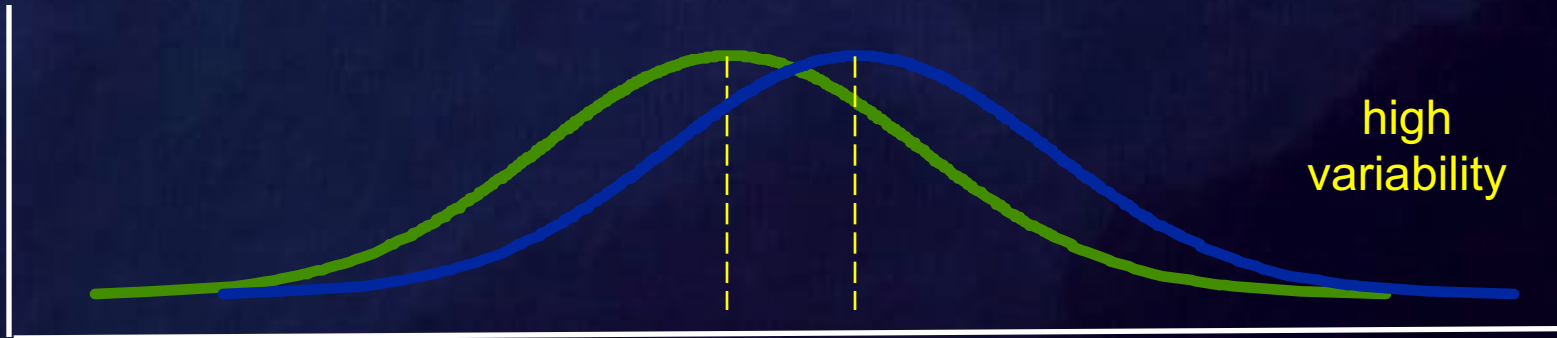
Is there a *difference*?
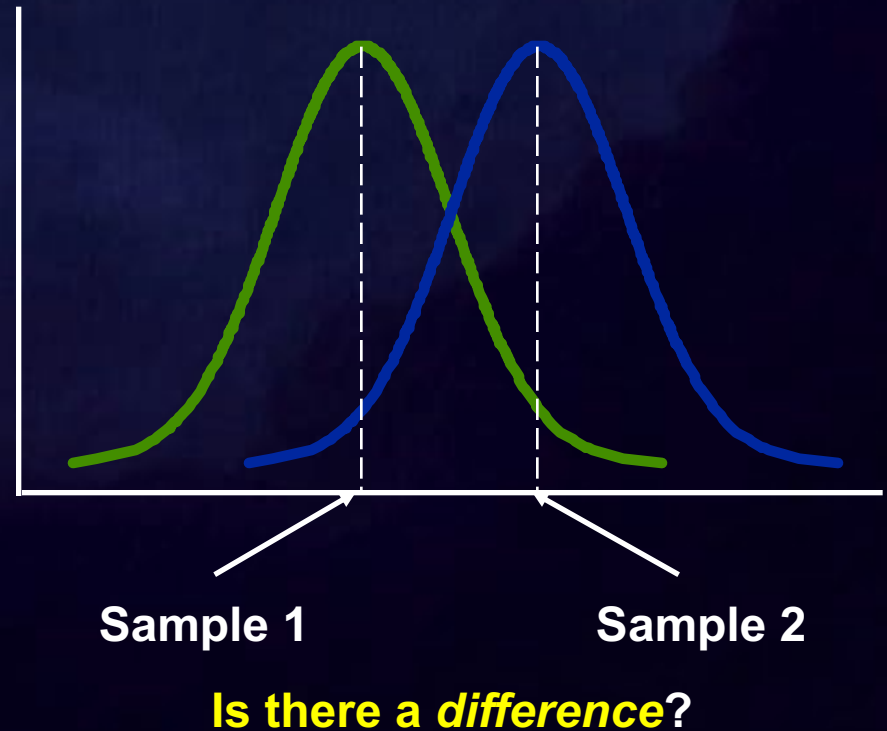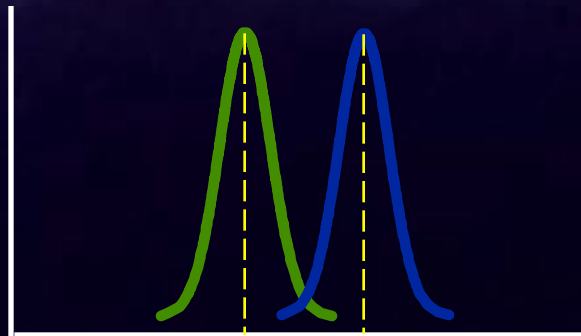
medium variability
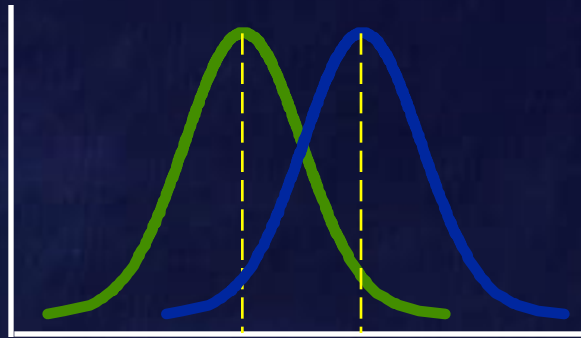
The mean difference is the *same* for all three cases

high variability

low variability

Differences

- Statistical difference is a function of the *difference between means* relative to the *variability*

- Small difference between means with large variability could be due to *chance*

- The t-test and the z-test essentially are *signal-to-noise* ratios
    - Large signal is easier to see difference between means
    - Large noise makes it difficult to see a significant difference

Sample 1   Sample 2

Is there a *difference*?

Data Analysis in Geography
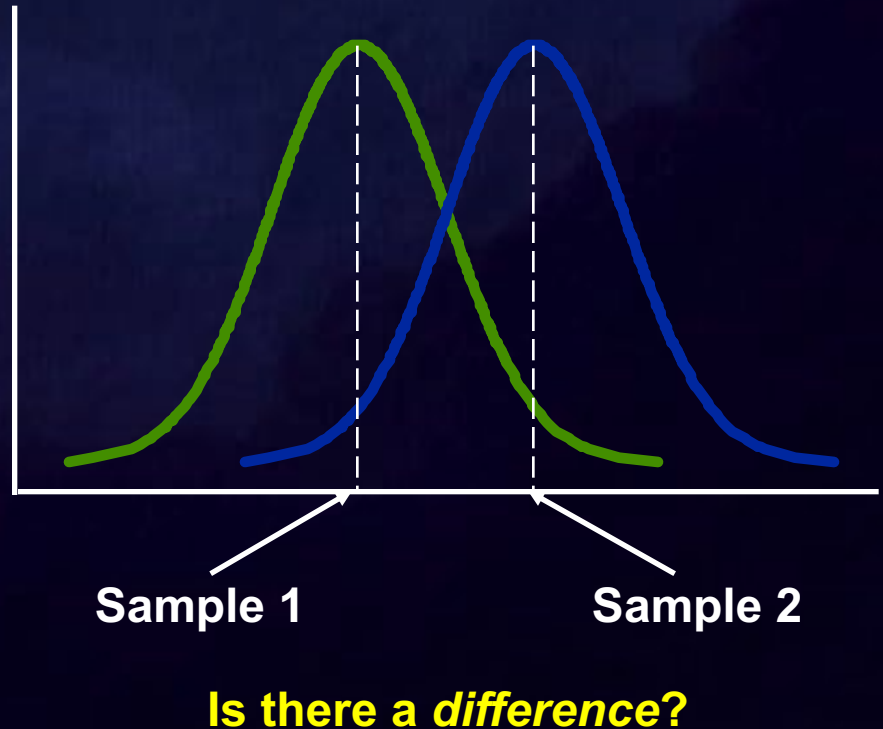
Hypothesis Testing

- **Research Question:** Is there a statistically significant difference between rainfall between Texas and Washington State?
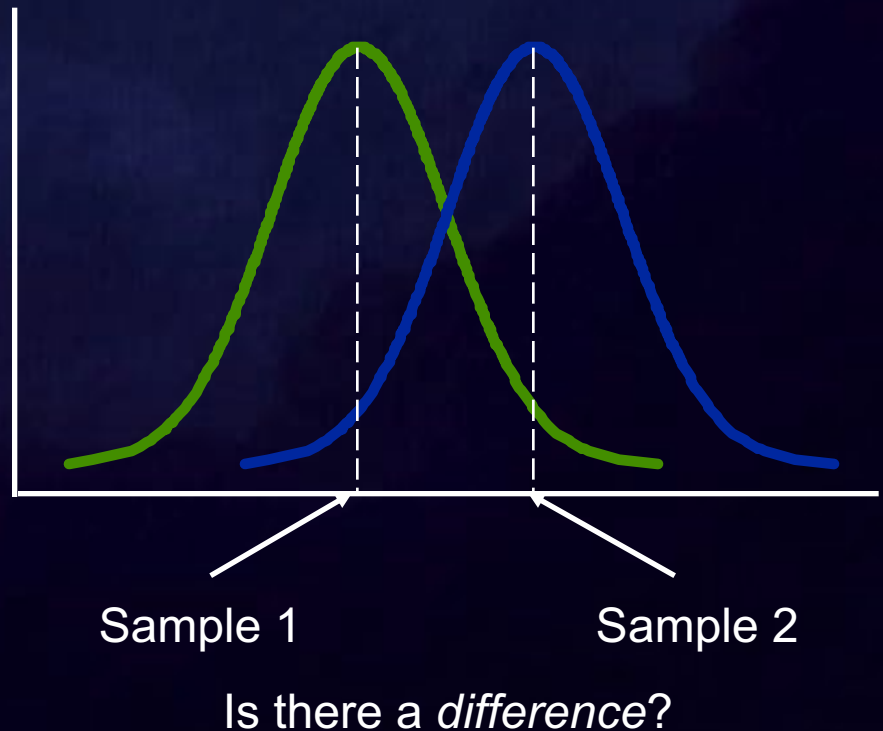
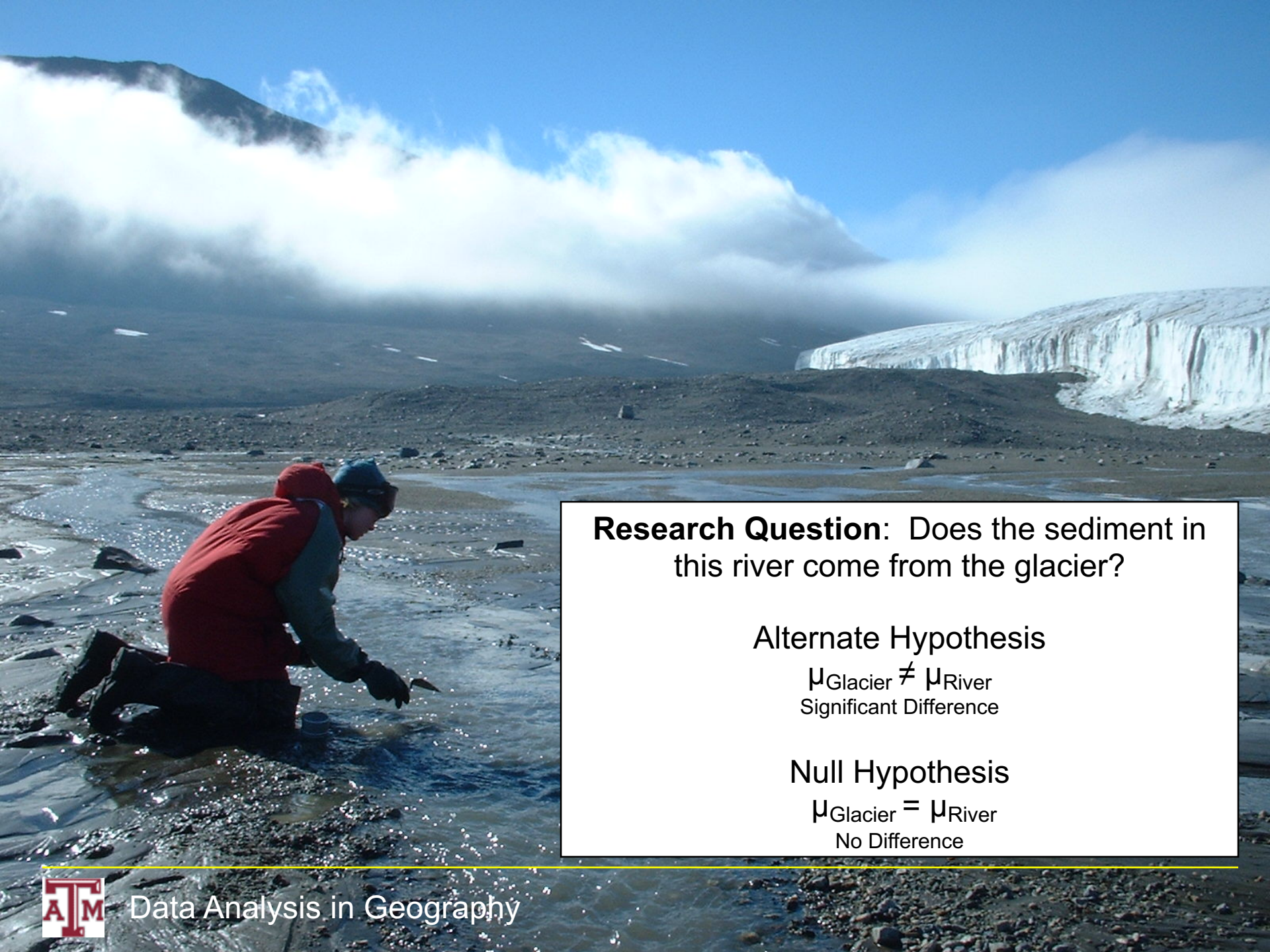### Alternate Hypothesis

$\mu_{Texas} \neq \mu_{Washington}$

Significant Difference

Null Hypothesis

$\mu_{Texas} = \mu_{Washington}$
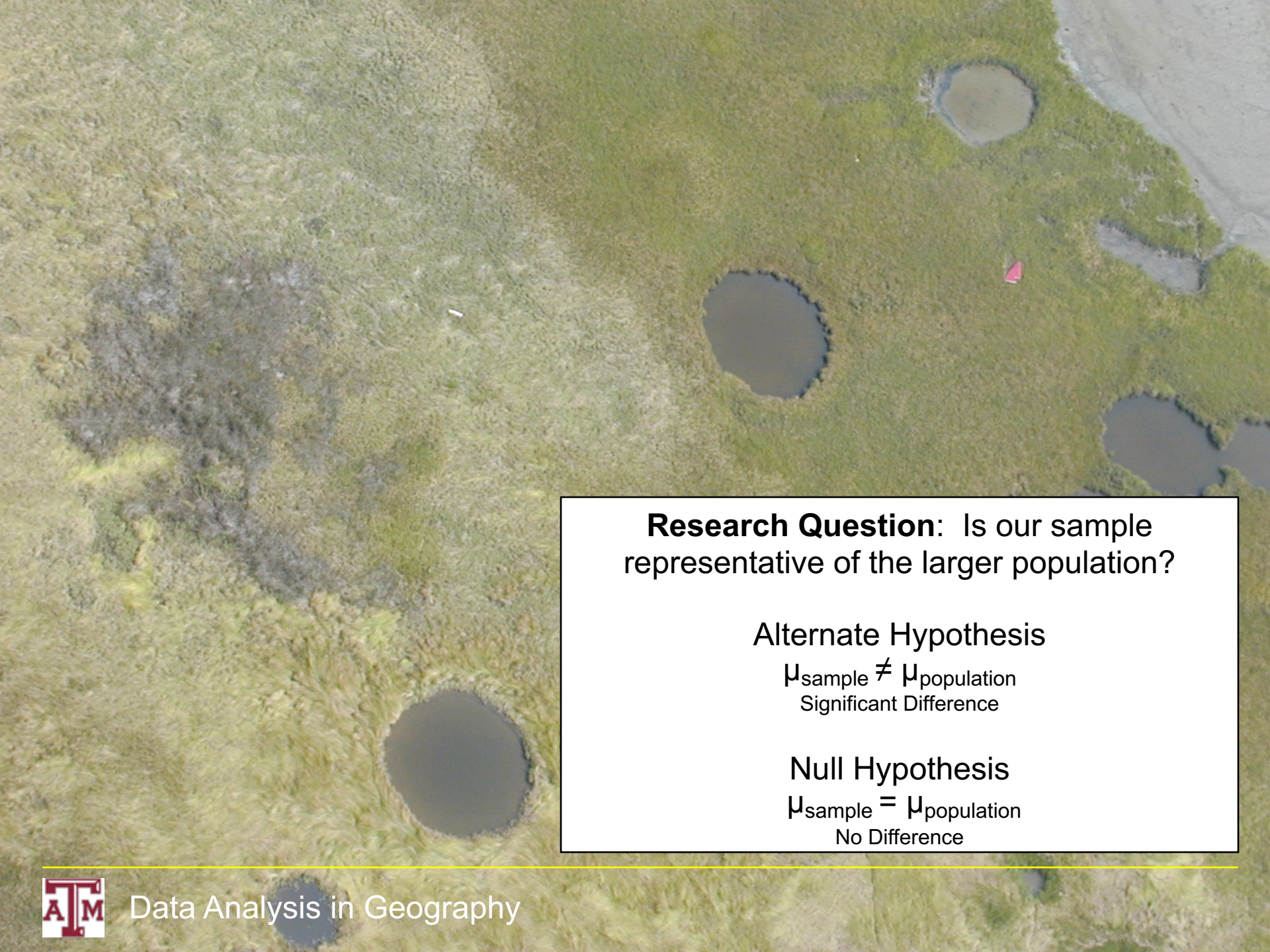
No Difference

Sample 1   Sample 2

Is there a *difference*?

Research Question: Does the sediment in this river come from the glacier?

Alternate Hypothesis

$\mu_{Glacier} \neq \mu_{River}$
Significant Difference

Null Hypothesis

$\mu_{Glacier} = \mu_{River}$
No Difference

Data Analysis in Geography

Research Question: Is our sample representative of the larger population?
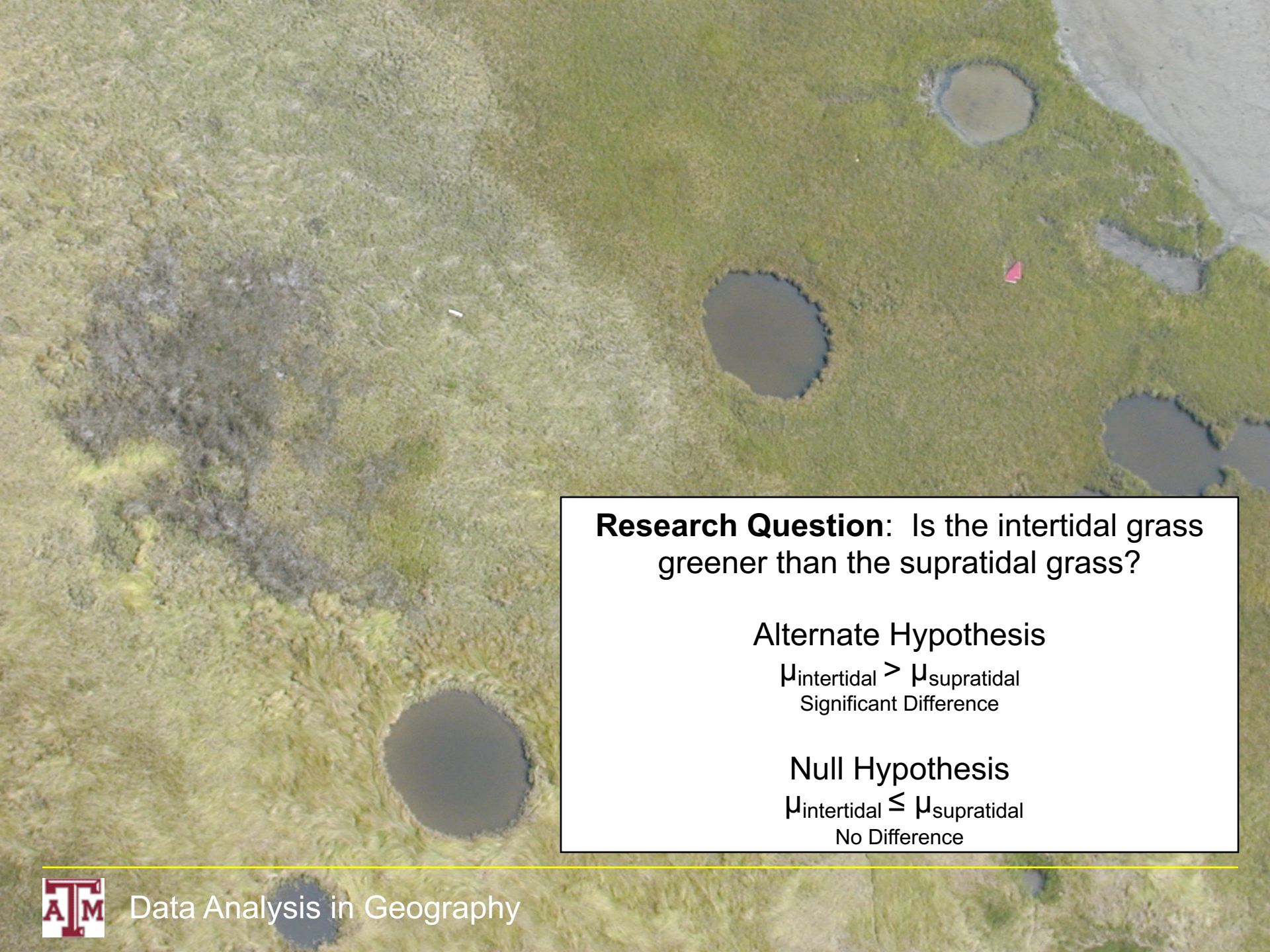
Alternate Hypothesis

$\mu_{sample} \neq \mu_{population}$
Significant Difference

Null Hypothesis

$\mu_{sample} = \mu_{population}$
No Difference

Research Question: Is the intertidal grass greener than the supratidal grass?

Alternate Hypothesis

$\mu_{intertidal} > \mu_{supratidal}$
Significant Difference

Null Hypothesis

$\mu_{intertidal} \leq \mu_{supratidal}$
No Difference

Data Analysis in Geography

Research Question: Is there a difference in the color of the grass?

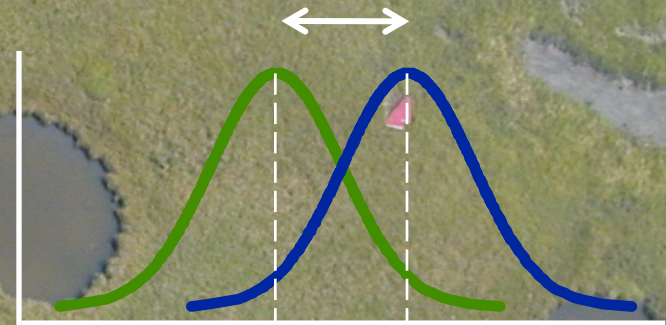Alternate Hypothesis

$\mu_{intertidal} \neq \mu_{supratidal}$

Significant Difference

Null Hypothesis

$\mu_{intertidal} = \mu_{supratidal}$

No Difference

Two-tailed Test

Research Question: Is the intertidal grass greener than the supratidal grass?

Alternate Hypothesis

$\mu_{intertidal} > \mu_{supratidal}$

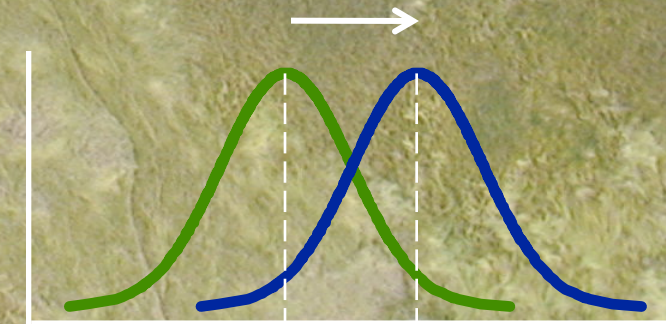Significant Difference

Null Hypothesis

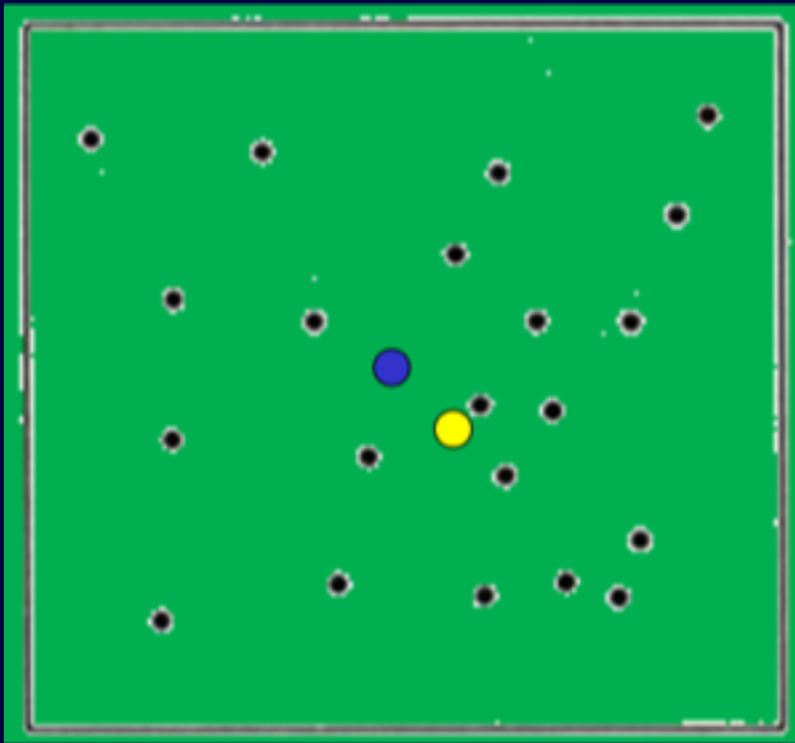$\mu_{intertidal} \leq \mu_{supratidal}$

No Difference

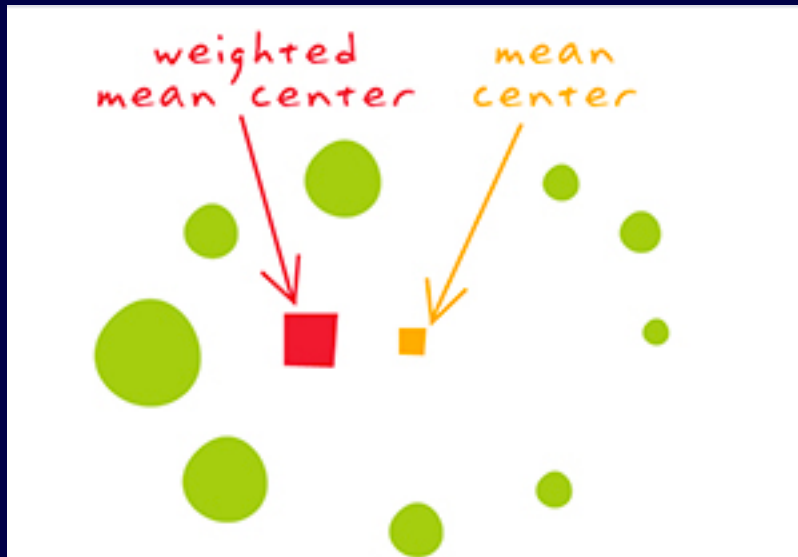**One-tailed Test**

Data Analysis in Geography

Mean Center

- Mean center (X_{bar}, Y_{bar}) is calculated as the arithmetic average of x and the average of y (mean center is not the same as the geometric center or centroid)

$$\overline{X} = \frac{x_1 + x_2 + ... + x_n}{n}$$

$$\overline{Y} = \frac{y_1 + y_2 + ... + y_n}{n}$$

Weighted Mean Center

Value dependent center

- **Weighted mean center** is weighted based on the sample values
- Previous method is "equal-weighted"

$$\overline{X} = \frac{v_1 x_1 + v_2 x_2 + \ldots + v_3 x_n}{v_1 + v_2 + \ldots + v_3}$$

$$\overline{Y} = \frac{v_1 y_1 + v_2 y_2 + \ldots + v_3 y_n}{v_1 + v_2 + \ldots + v_3}$$

Data Analysis in Geography

Let's open up R…

Lecture 8:
Comparing Distributions