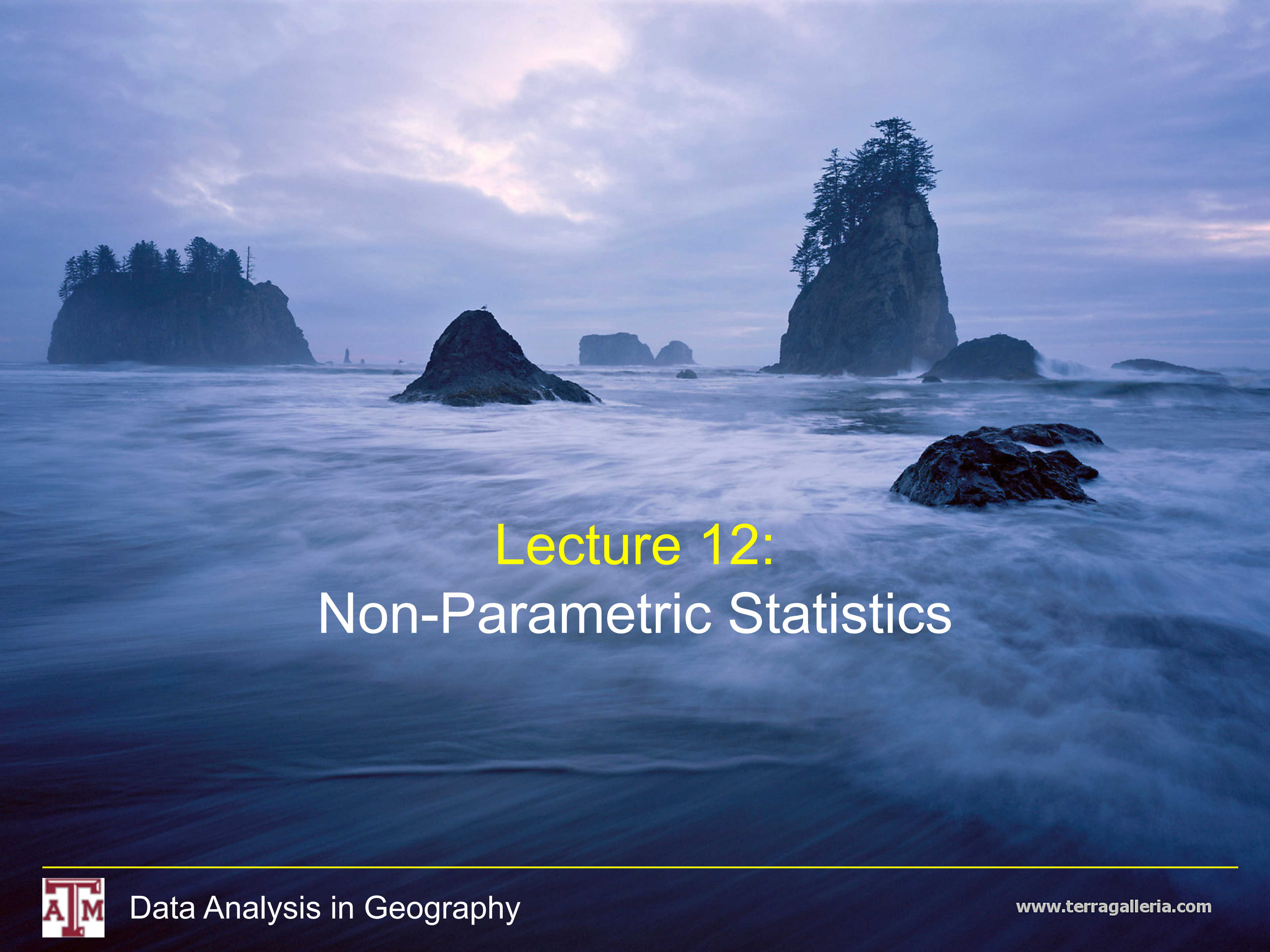


Lecture 12:
Non-Parametric Statistics

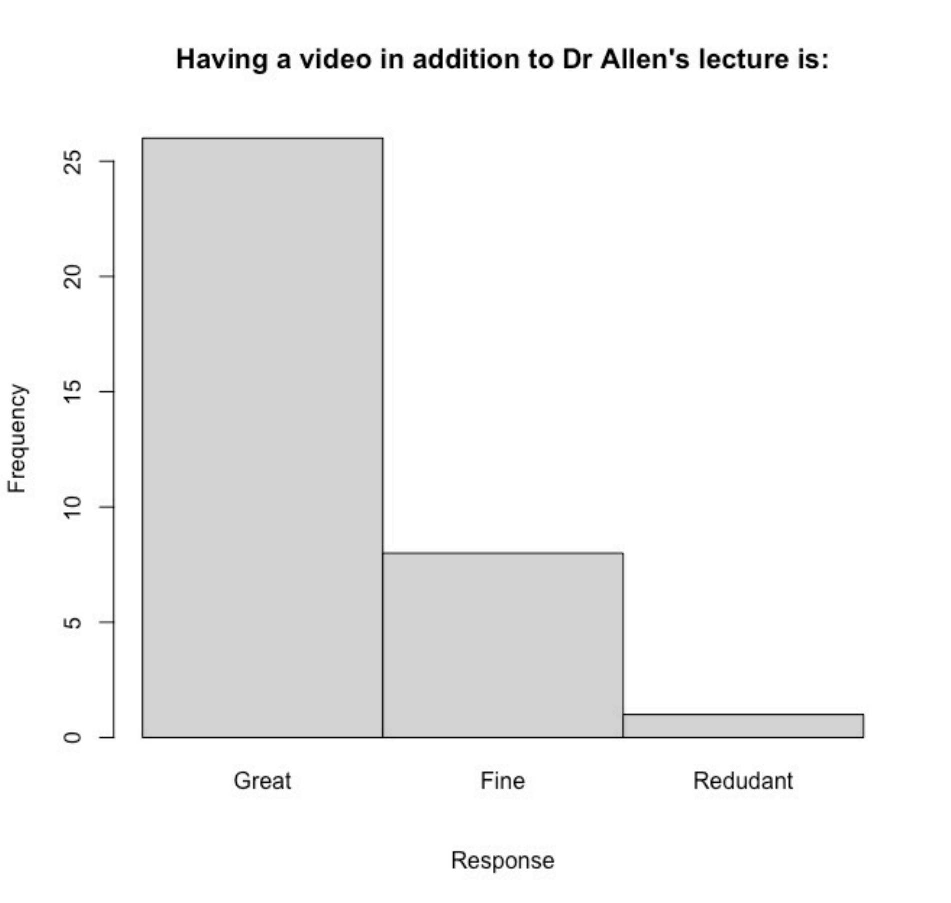
www.terragalleria.com

Announcements

Extension Granted: Assignment #3 is due tomorrow at 3:55 pm

Assignment #4 will be assigned next class

Survey Results



- Z-test video: https://www.youtube.com/watch?v=L8QR7wxmmQg

Importance of Statistics

1. How to effectively collect data and the types of data
2. Using descriptive statistics to assess our data
3. Expressing probability using collected samples
4. Detecting difference in means of two samples
   - Sample vs. population
   - Sample vs. sample
5. Detecting differences in means of more than two samples
   - ANOVA

Our descriptions and tests have mostly assumed normal distributions

Importance of Statistics

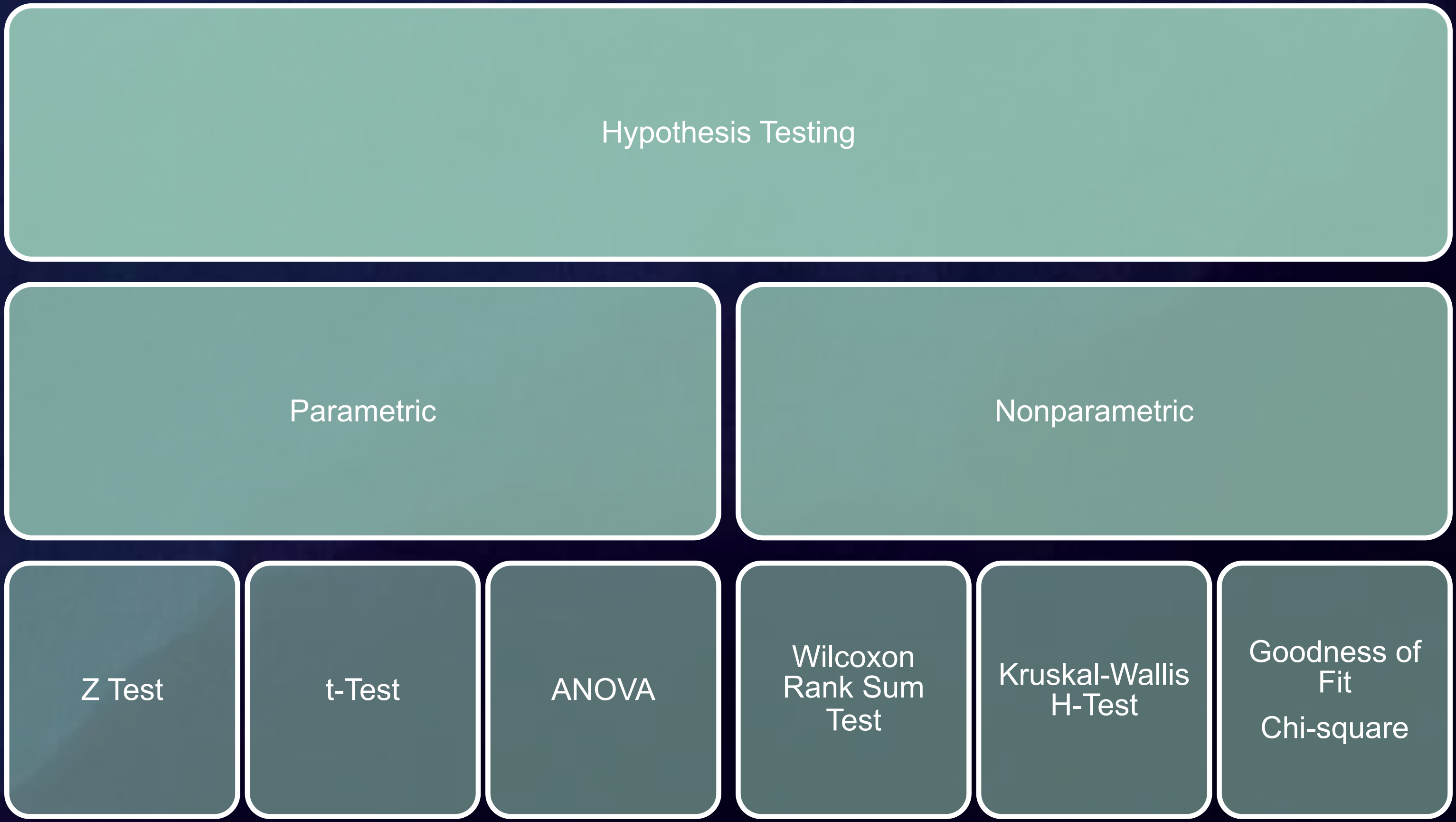
- The ability to collect and analyze quantitative data is one of the basic and fundamental tools for a professional geographer.
- Often, the expression of a study’s results using proper statistics is the most important deciding factor as to whether the methods are accepted.

Importance of Statistics

1. To describe and summarize spatial data.
2. To estimate the probability of outcomes for an event at a given location.
3. To use samples of geographic data to infer characteristics for a larger set of geographic data (population).
4. To determine if the magnitude or frequency of some phenomenon differs from one location to another.
5. To learn whether an actual spatial pattern matches some expected pattern.

https://en.wikipedia.org/wiki/Statistical_geography

Non-Parametric Tests



Many More Tests Exist!

Parametric Tests

- Normality Assumption: The two populations are assumed to both follow a normal distribution
- Data Type Assumption: Interval or ratio data
- Variance Assumption: All samples have approximately the same variance
- Population variance: Z test requires that the population variance is well approximated by the sample variance

Non-Parametric Tests

- Normality Assumption: Don't need to have normal distributions
- Data Type Assumption: Data can be measured on any scale
- Variance Assumption: No strict assumptions about the nature of the data
- BUT are more difficult for large samples and you lose information when data converted to ordinal or nominal

Non-Parametric Tests

Pros:

- Normality Assumption: Don’t need to have normal distributions
- Data Type Assumption: Data can be measured on any scale
- Variance Assumption: No strict assumptions about the nature of the data
- Population Variance: No information about the population is needed

Non-Parametric Tests

Cons:

- More difficult for large samples
- Lose information when data converted to categorical data (e.g., ordinal or nominal)
- Difficult to compute by hand for Large Samples
- Lookup tables are not readily available

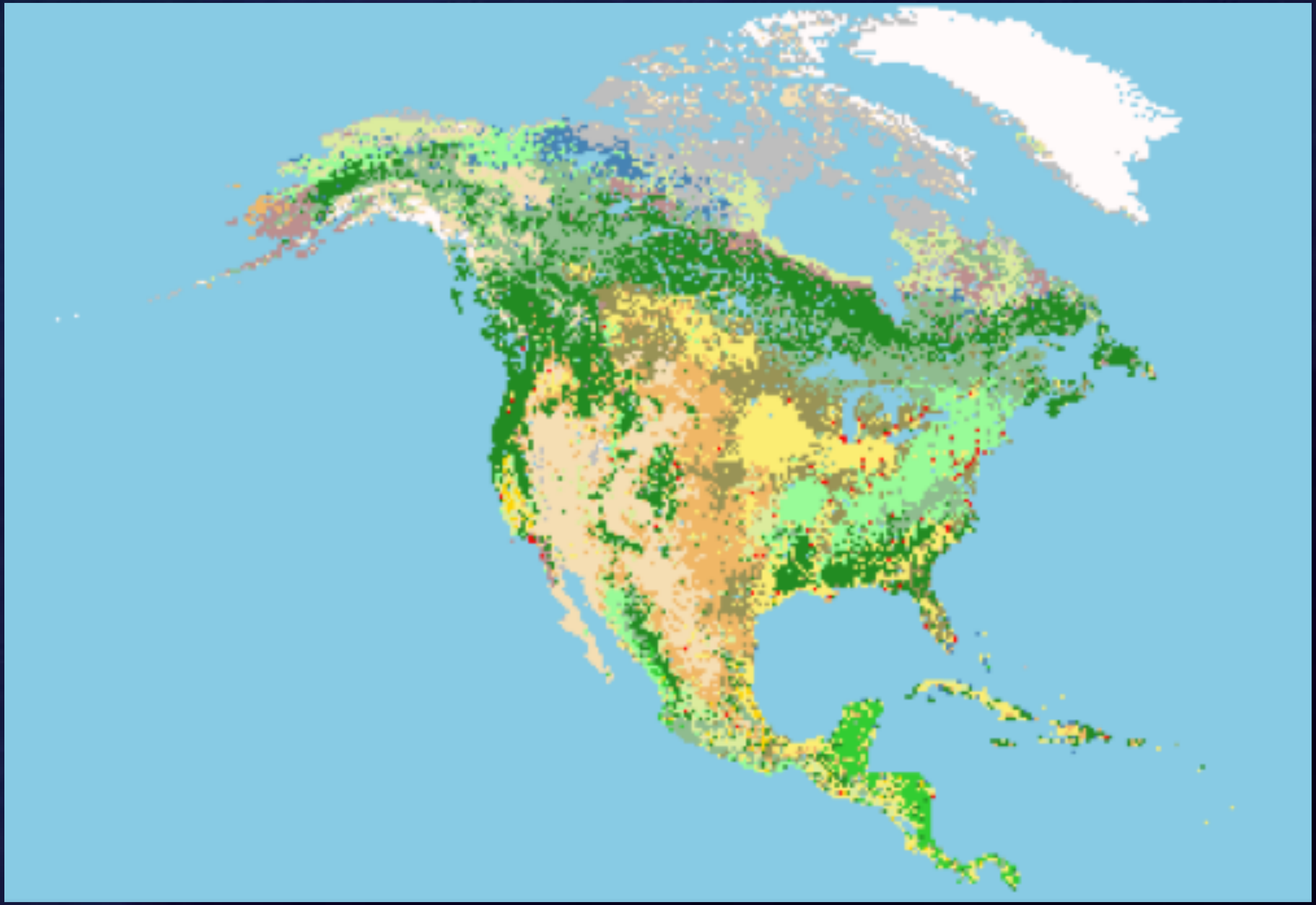
When to use Non-Parametric Tests

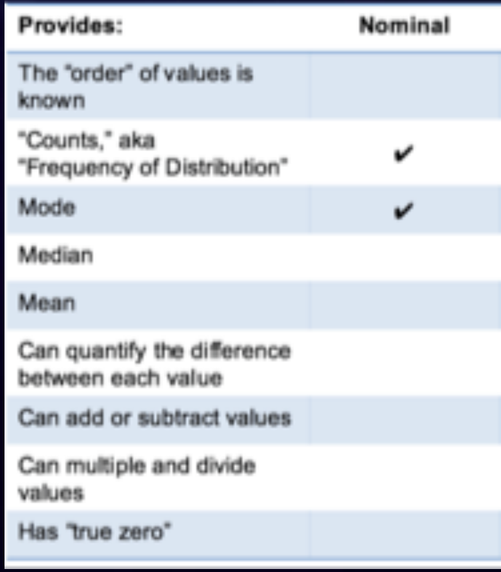
In a nutshell: When it is clear your dataset does not have a normal distribution

Specific reasons:

- When the data are nominal or ordinal
- When the data are rankings
- When the dataset suffers from outliers
- When the observed quantity is difficult to detect

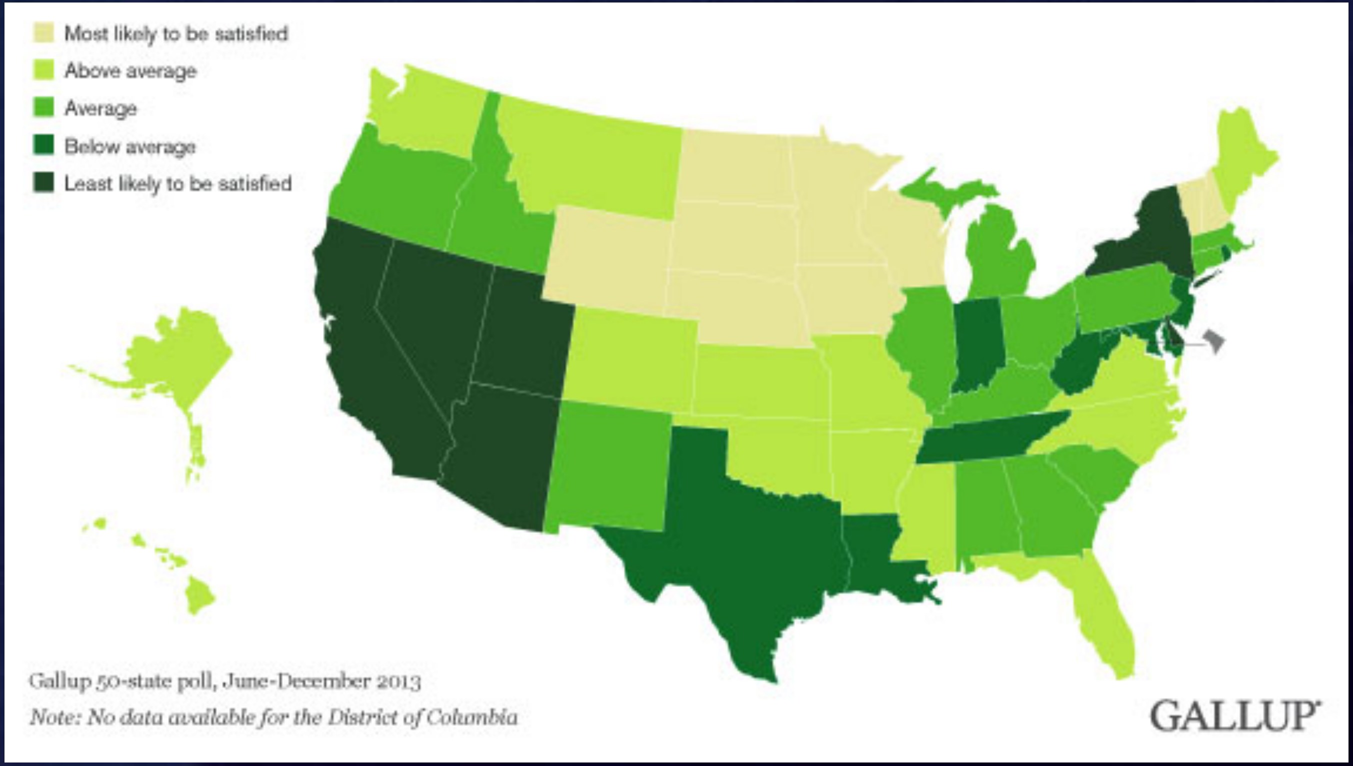
Nominal Data





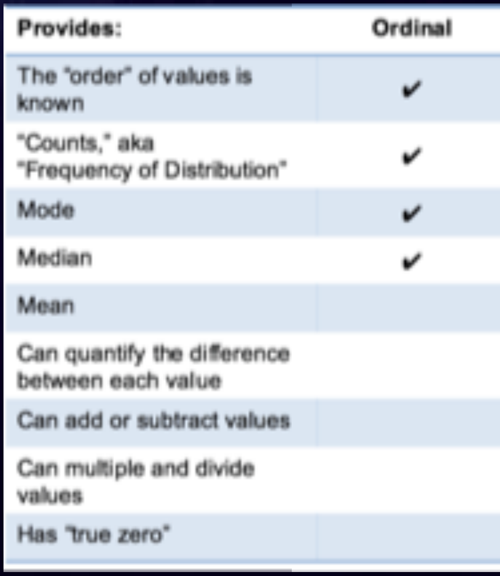
Provides:	Nominal
The "order" of values is known	
"Counts," aka "Frequency of Distribution"	✔
Mode	✔
Median	
Mean	
Can quantify the difference between each value	
Can add or subtract values	
Can multiple and divide values	
Has "true zero"	

Ordinal Data



Gallup 50-state poll, June-December 2013

Note: No data available for the District of Columbia

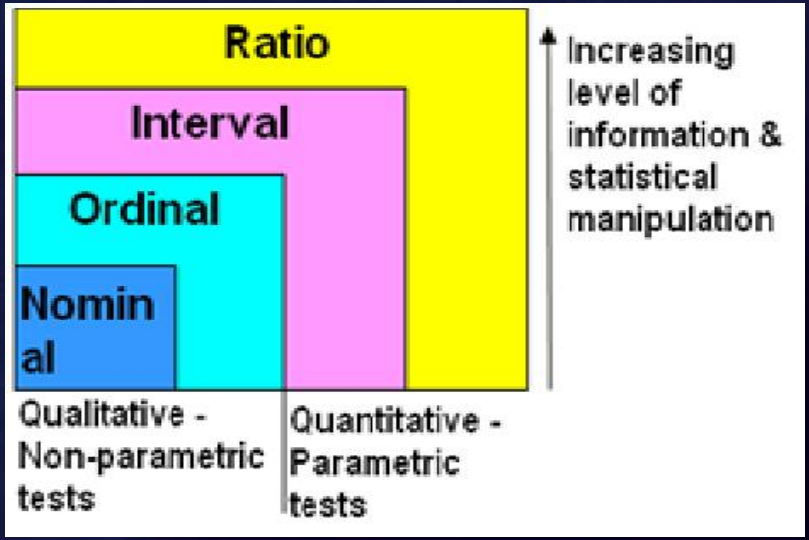


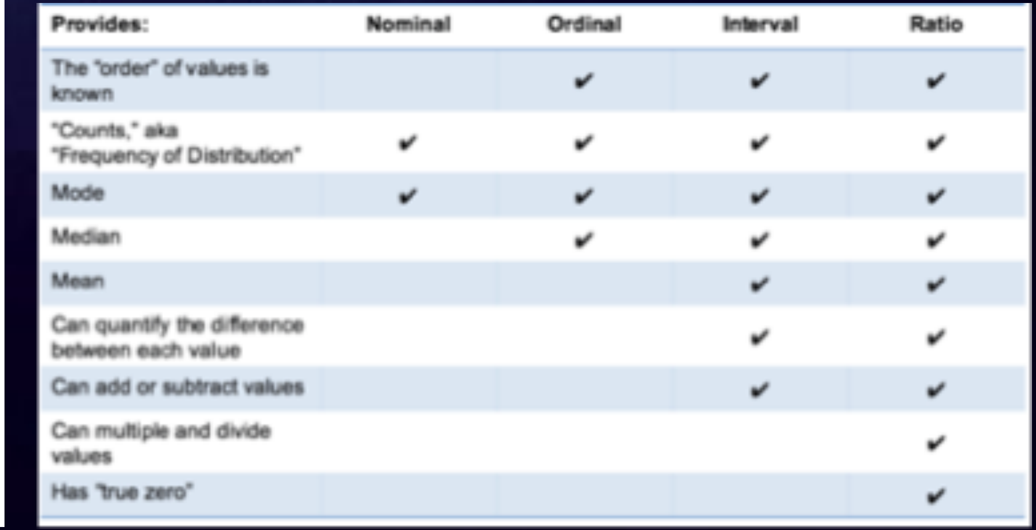
Provides:	Ordinal
The "order" of values is known	✔
"Counts," aka "Frequency of Distribution"	✔
Mode	✔
Median	✔
Mean	
Can quantify the difference between each value	
Can add or subtract values	
Can multiple and divide values	
Has "true zero"	

When to use Non-Parametric Tests

- When it is clear your dataset does not have a normal distribution

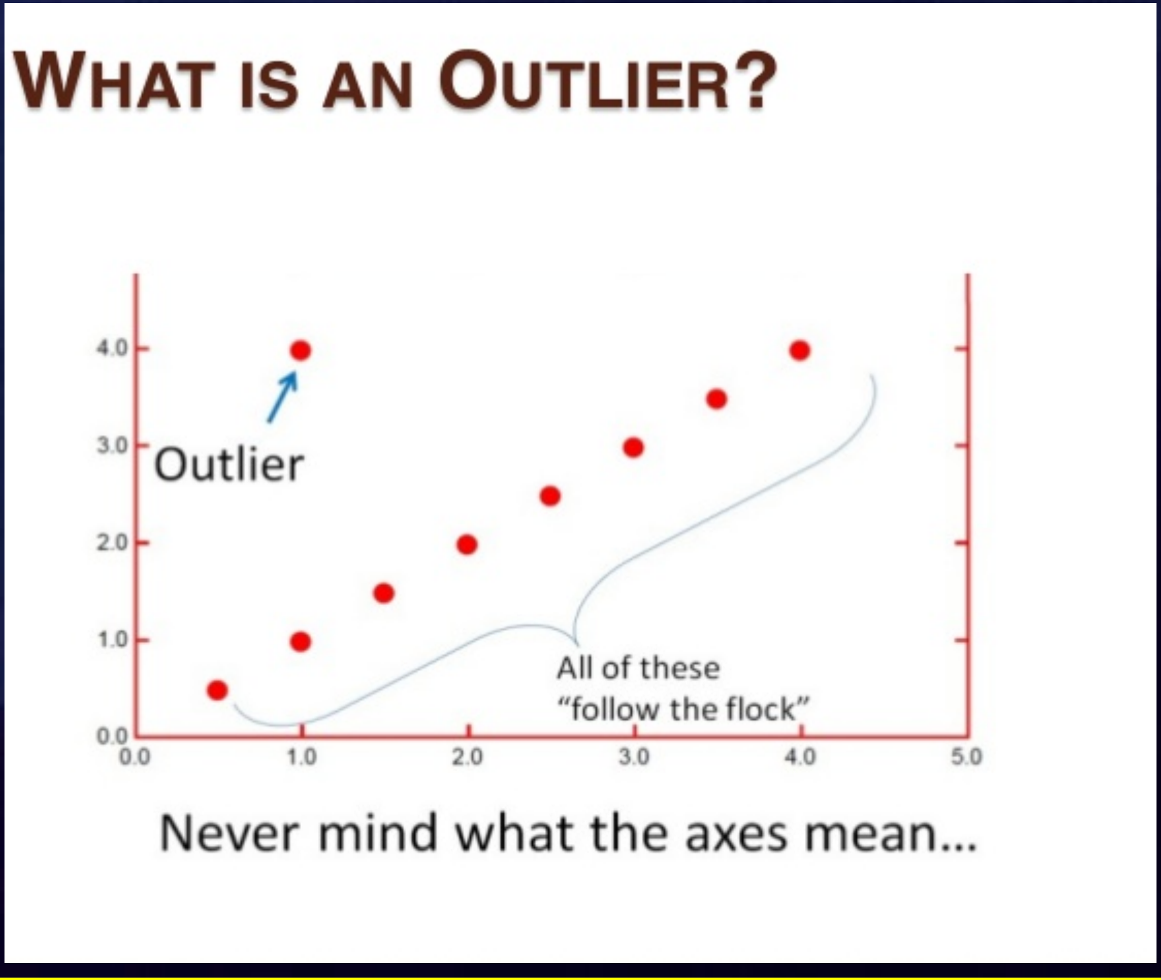
- Relatively normal distribution contains outliers


Ratio
Interval
Ordinal
Nominal
Increasing level of information & statistical manipulation
Qualitative - Non-parametric tests
Quantitative - Parametric tests



Provides:	Nominal	Ordinal	Interval	Ratio
The "order" of values is known		✓	✓	✓
"Counts," aka "Frequency of Distribution"	✓	✓	✓	✓
Mode	✓	✓	✓	✓
Median		✓	✓	✓
Mean			✓	✓
Can quantify the difference between each value			✓	✓
Can add or subtract values			✓	✓
Can multiple and divide values				✓
Has "true zero"				✓

Outliers



WHAT IS AN OUTLIER?

Never mind what the axes mean...

Loss of Information

Two schools of thought:

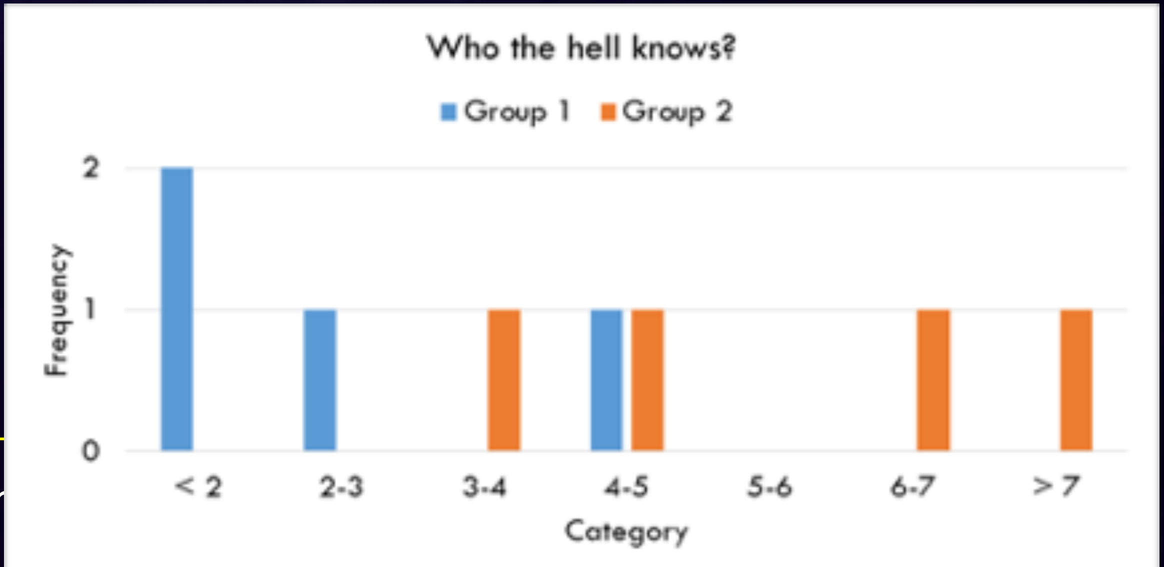
- Pro-parametric
  - Because information is discarded, nonparametric procedures can never be as powerful when parametric tests can be used
  - Use parametric testing if a distribution is normal!

- Pro-non-parametric
  - There are too many assumptions needed for parametric tests, which usually are not met in real experiments
  - Use non-parametric testing unless there is strong and compelling evidence that the distribution of errors is normal!

Loss of Information

Data is ranked… (we will get more into this next lecture)

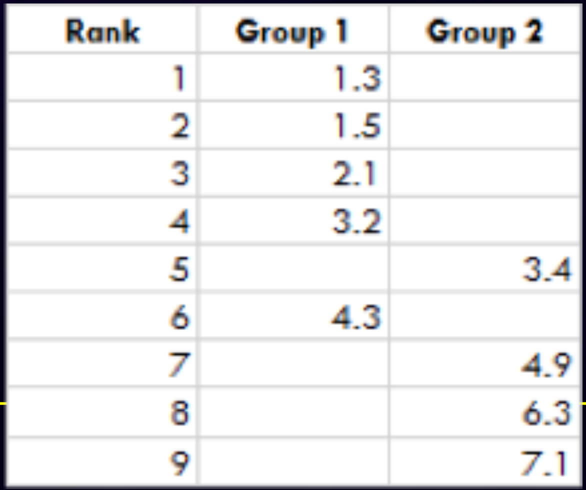
- Example: We have observed values:
  - Group 1: 3.4, 4.9, 6.3, 7.1
  - Group 2: 1.3, 2.1, 1.5, 4.3, 3.2
  - **Are the groups significantly different?**



Loss of Information

Data is ranked…

- Example: We have observed values:
  - Group 1: 3.4, 4.9, 6.3, 7.1
  - Group 2: 1.3, 2.1, 1.5, 4.3, 3.2
  - **Are the groups significantly different?**



Rank	Group 1	Group 2
1	1.3	
2	1.5	
3	2.1	
4	3.2	
5		3.4
6	4.3	
7		4.9
8		6.3
9		7.1

Remarks on Non-parametric Statistics

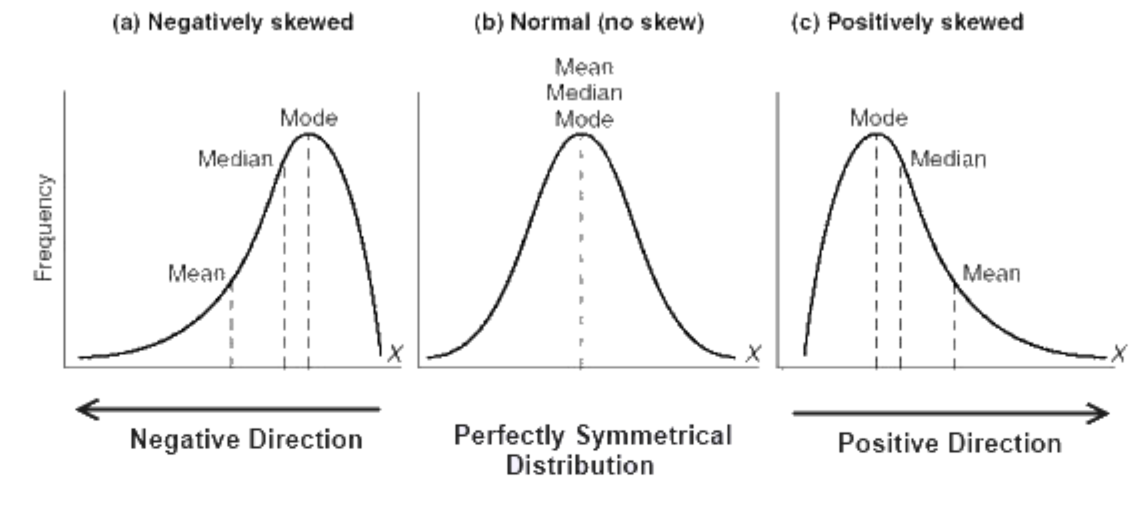
- **Fewer assumptions**
  - Pro: More confidence
  - Con: More general null hypothesis
- **Good choice when normality of the data cannot be assumed**
  - Pro: Reject the null hypothesis with a non-parametric test = pretty sure that your samples are different
  - Con: Much more difficult to get significant conclusions

Remarks on Non-parametric Statistics

- **‘Distribution-free’ may be more appropriate**
  - Parametric tests/distributions have fixed numbers of parameters
  - Non-parametric has more parameters as sample(s) grow
  - Explains why often difficult to compute by hand
- **Most non-parametric tests about the population center are tests about the median instead of the mean**
  - Requires you to modify the null hypotheses
  - Does not answer the same question as the corresponding parametric procedure

Remarks on Non-parametric Statistics

- **‘Distribution-free’ may be more appropriate**
  - Parametric tests/distributions have fixed numbers of parameters
  - Non-parametric has more parameters as sample(s) grow
  - Explains why often difficult to compute by hand

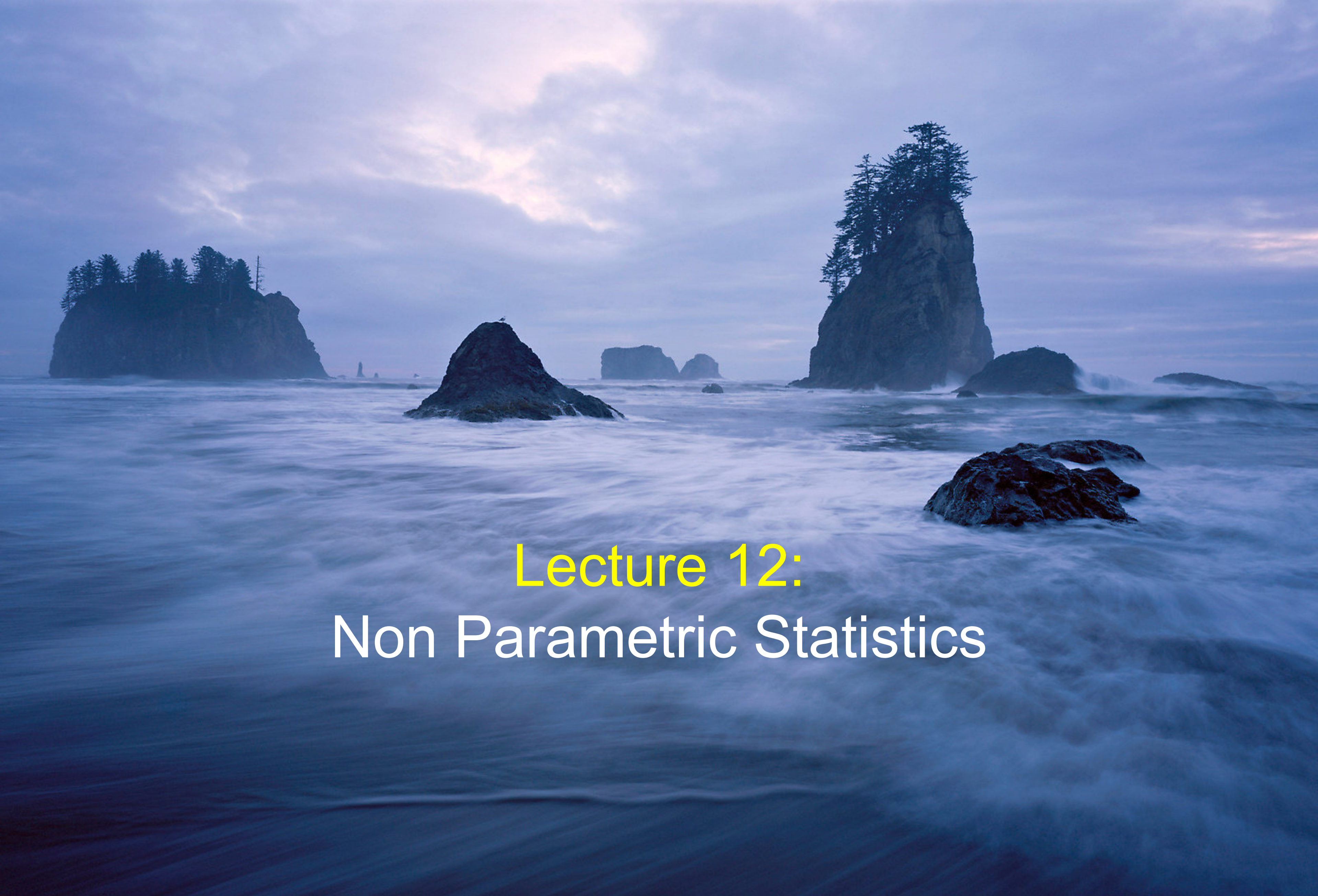


Supplemental Reading

- **Good explanation about the differences between and appropriateness of parametric and non-parametric tests:**

- **Look in Course Materials in eCampus:**

Colquhoun, D. (1971). *Lectures on biostatistics: an introduction to statistics with applications in biology and medicine.* David Colquhoun.



Lecture 12:
Non Parametric Statistics

www.terragalleria.com