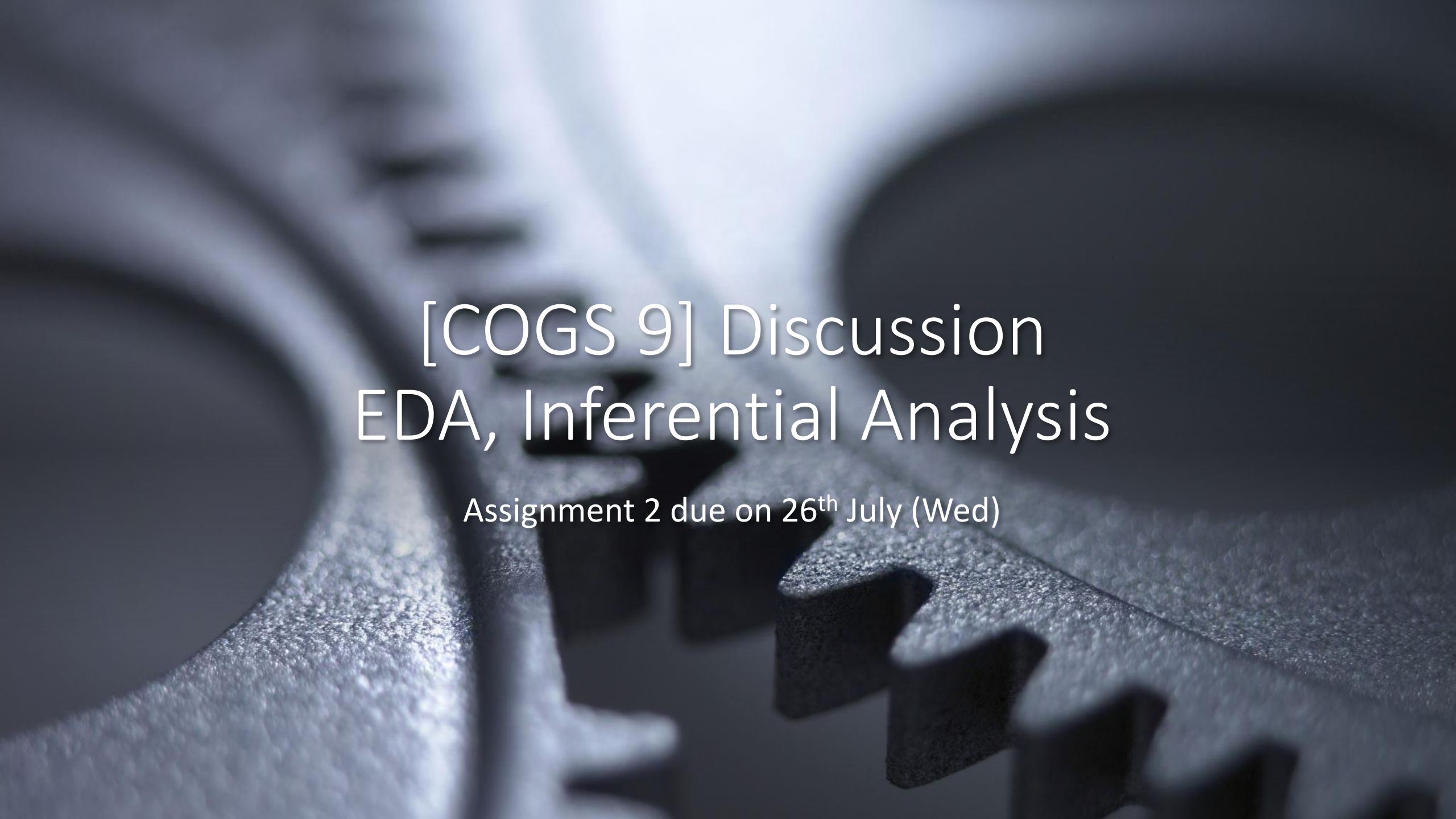


[COGS 9] Discussion
EDA, Inferential Analysis

Assignment 2 due on 26th July (Wed)

Exploratory Data Analysis

Data -> Descriptive Analysis -> Exploratory Analysis -> Product

Exploratory: Inferential, Predictive, Causal, Mechanistic

- Inferential: Statistics, Frequentist, Bayesian, Text & Geospatial analysis
- Predictive: Statistical Learning/ML, Deep, Reinforcement Learning
- Causal: How variable X correlates to Y
- Mechanistic: How much does variable X affect Y

Univariate, Bivariate, Multivariate

Explanatory (Independent) vs Response (Dependent) variables

Source of data (Zipcode vs hometown), explore missing data

Don’t do EDA to give you the result you want

Exploratory Data Analysis

Checklist of things to do during EDA

- ❑ Investigate missing values
- ❑ Understand outliers
- ❑ Add filters, transform and scale data
- ❑ Calculate numerical summaries
- ❑ Generate plots to explore relationships
- ❑ Handle proportions correctly
- ❑ Use tables to scan data
- ❑ Search for patterns

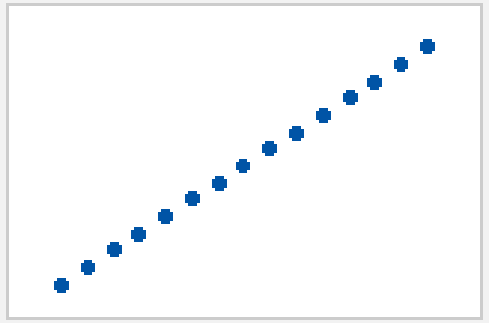
EDA
Demo

Correlation Coefficients

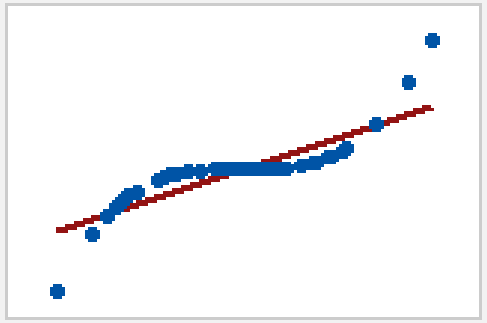
Pearson Correlation = Linear relationship between two variables

Spearman Correlation = Monotonic relationship between two sets

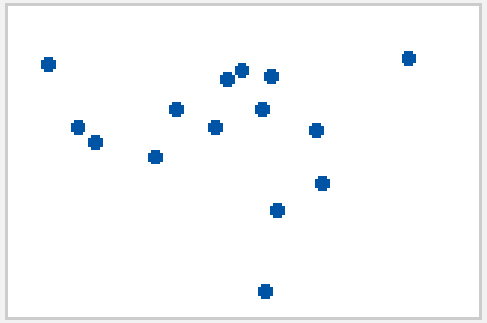
Pearson = +1, Spearman = +1



Pearson = +0.851, Spearman = +1



Pearson = −0.093, Spearman = −0.093



T-Test

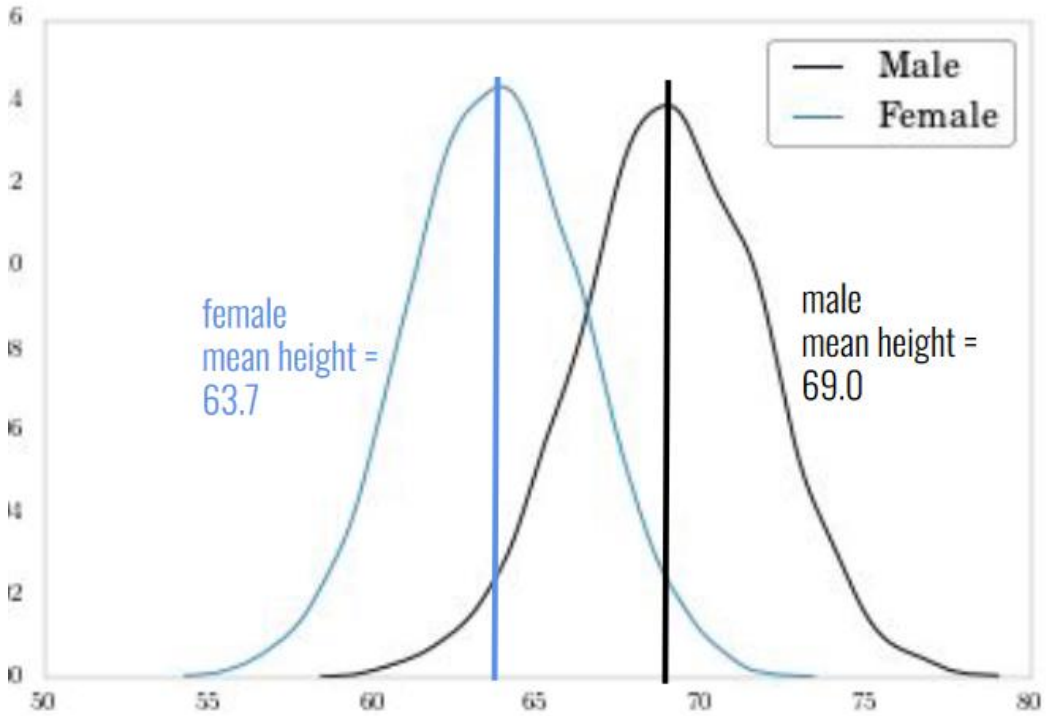
Test for differentiating between the means of 2 groups

- Data should be continuous
- Normally distributed
- Large enough sample size
- Equal variance between groups

Greater magnitude of T implies that there is a statistically significant difference b/w the 2 groups

t-statistic: -95.6

p-value < 0.001



Inferential Analysis Demo