# STAT 216: Course Outline

## Chapter 1: Introduction to data

1. Martian Alphabet
   - observational units
   - types of variables (categorical/quantitative)
   - proportion
   - frequency and relative frequency bar plots
   - distribution
2. American Indian Address + Lab
   - sample vs population
   - sampling bias (selection, non-response, response)
   - (simple) random sample
   - explanatory/response variable
   - confounding variable
   - observational study
   - randomized experiment
     - control, randomization, replication, blocking
   - scope of inference
     - generalize
     - causation

## Chapter 2: Exploratory data analysis

3. Categorical EDA
   - two-way (contingency) tables
   - independence (of two variables)
   - distribution
   - summary measures:
     - proportion (unconditional, conditional, row, column)
     - **difference in proportions**
     - **relative risk**
   - plots:
     - bar plots for one categorical variable
     - segmented bar plots for two categorical variables
     - mosaic plots
   - Simpson's Paradox (book only)
4. IMDb Movie Reviews - one quantitative variable
   - summary measures:
     - center: mean, median
     - variability/spread: interquartile range (IQR), standard deviation
     - percentiles, quartiles
     - five-number summary
   - shape: symmetric, right/positive skew, left/negative skew
   - plots:
     - dot plot, histogram, density plot, box plot for one quantitative variable

5. Movie Budgets + Lab - one quantitative variable + one categorical variable
    - outliers and robust statistics
    - plots:
        - side-by-side boxplots or stacked dotplots/histograms/density plots for one categorical variable and one quantitative variable

## Chapter 3: Correlation and regression & Chapter 4: Multivariable models

6. Movie Profits + Lab - two quantitative variables + multivariable
    - least-squares regression line
    - summary measures:
        - correlation coefficient ($r$)
        - coefficient of determination ($R^2$)
        - $y$-intercept of regression line ($b_0$)
        - slope of regression line ($b_1$)
    - plots:
        - scatterplot for two quantitative variables (form, strength, direction, outliers)
        - scatterplot with categorical variable

## Chapter 5: Inference for categorical data

7. Helper/Hinderer + Lab - simulation-based inference for a single proportion
    - assumptions for inference
        - representative sample
        - independence (of observational units)
    - **statistic vs parameter**
    - **point estimate**
    - confidence interval
        - bootstrapping
        - confidence level
    - hypothesis testing
        - simulation-based
        - null and alternative hypotheses
        - null value of a parameter
        - null distribution of a statistic
        - p-value
        - significance level
        - statistical significance
8. Handedness of Male Boxers + Lab - theory-based inference for a single proportion
    - normal distribution and standard normal distribution
    - Z-score
    - 68-95-99.7 rule
    - Central Limit Theorem
    - sampling distribution of a statistic
        - standard deviation of a statistic
        - standard error of a statistic
    - confidence interval
        - theory-based
        - margin of error
        - confidence level

- hypothesis testing
  - theory-based
  - one-sided vs two-sided alternative hypothesis
  - test statistic
  - decision
  - Type 1 and Type 2 decision errors
  - power
  - statistical significance vs practical importance

9. The Good Samaritan + Lab - simulation-based inference for two categorical variables
10. Head Injuries and Helmet Use + Lab - theory-based inference for two categorical variables

## Chapter 6: Inference for quantitative data

11. COVID-19 and Air Pollution - paired mean difference
12. Construction Costs + Lab
13. Weather Patterns and Snowfall - difference in means
14. Homeless Housing + Lab

## Chapter 7: Inference for regression

15. Real Estate - simulation-based
    - **linear model ($\beta_0$, $\beta_1$, $\epsilon_i$)**
    - assumptions for inference
      - independence (of observational units)
      - linearity
16. Moneyball + Lab - theory-based
    - additional assumptions for inference
      - normality of errors
      - constant variance
    - residual plots:
      - residuals vs fitted values
      - histogram of residuals