Nathan Hill, Dylan Li, Rachel Roggenkemper, Kirina Sirohi

NFL PROJECTIONS AND PATTERNS

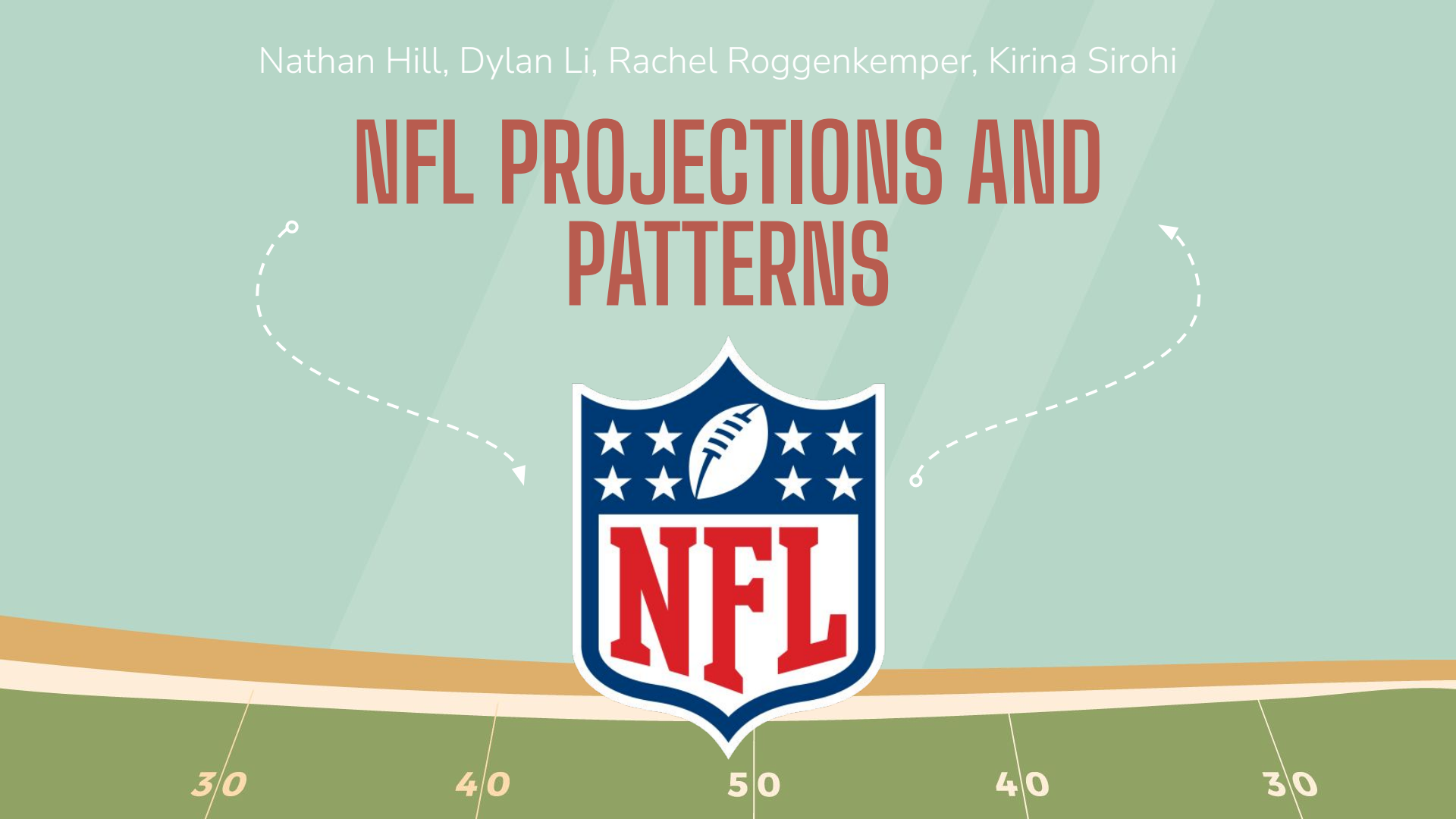
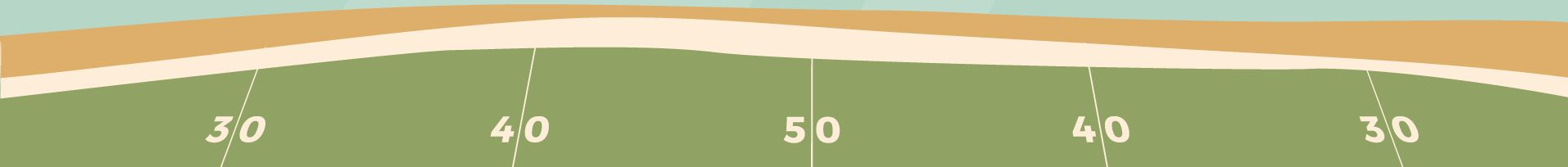
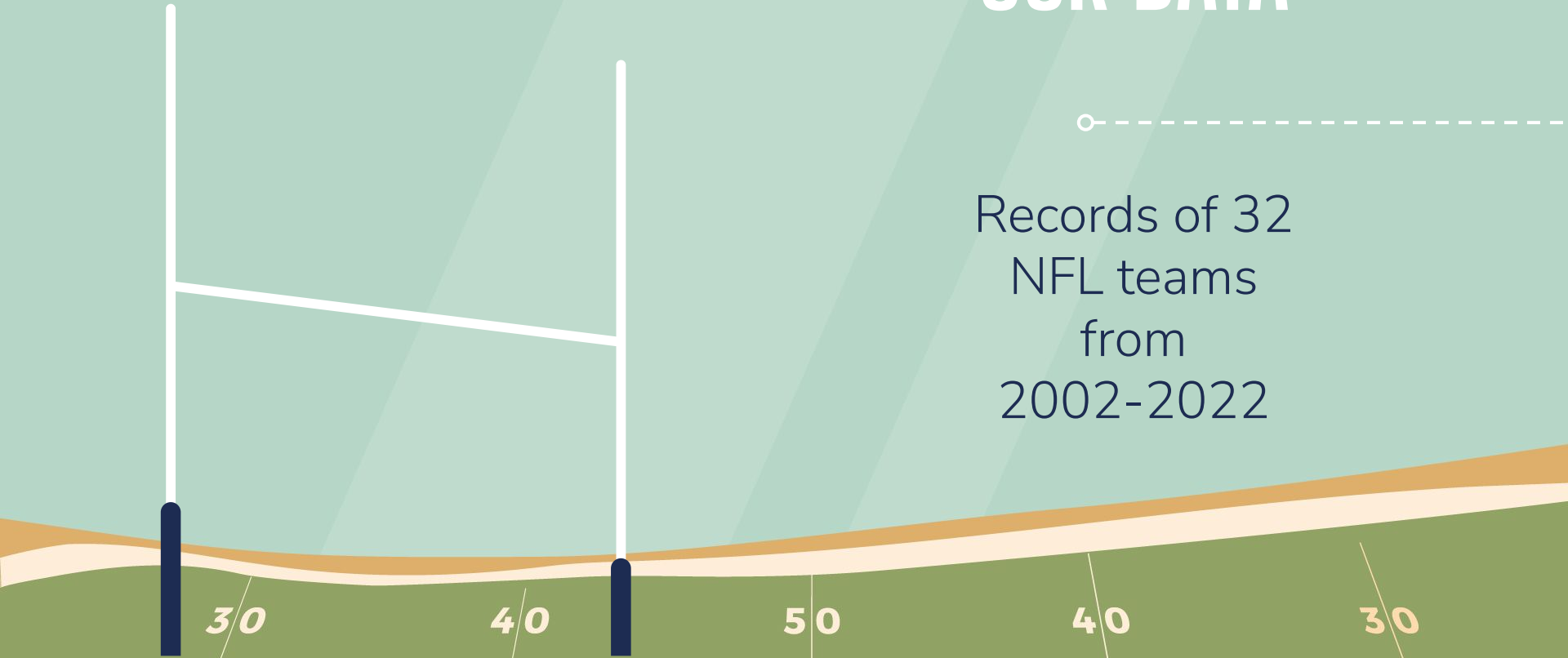


TABLE OF CONTENTS



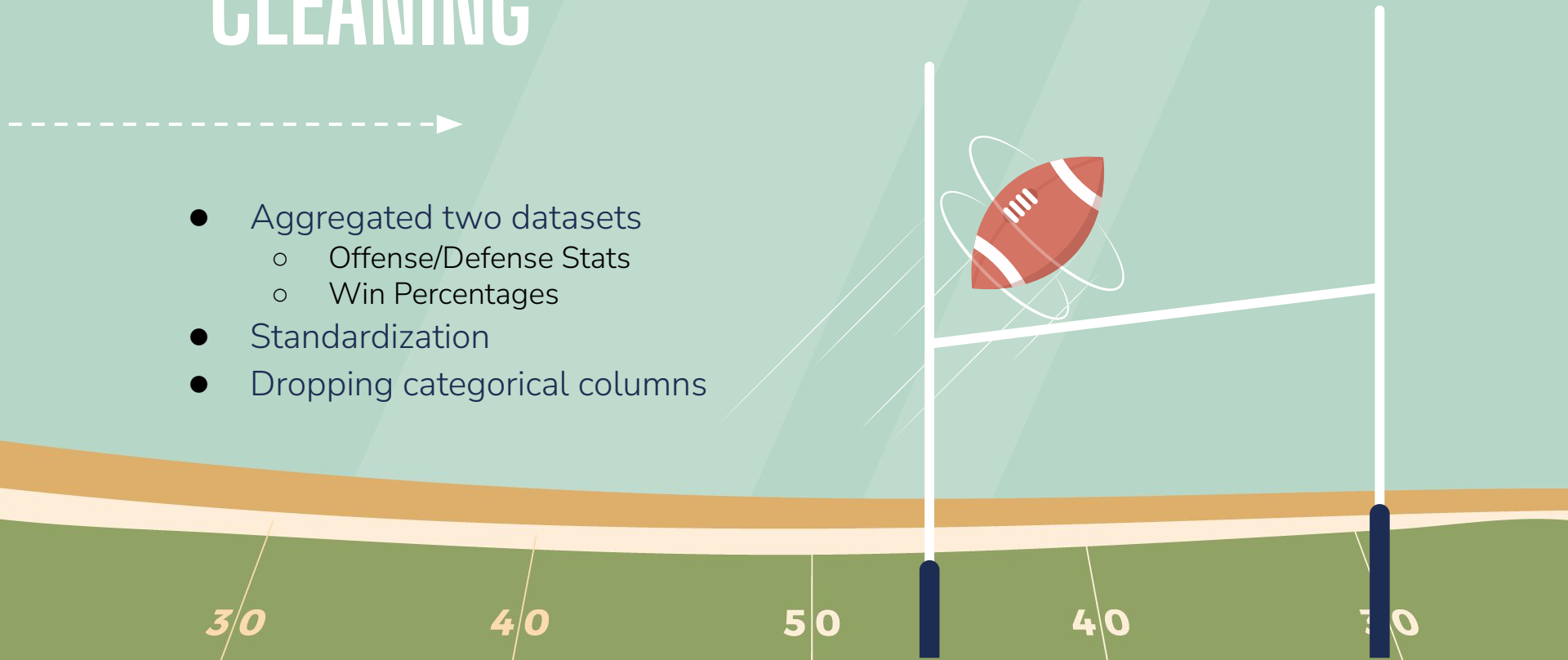
OUR DATA

Records of 32
NFL teams
from
2002-2022



CLEANING

- Aggregated two datasets
  - Offense/Defense Stats
  - Win Percentages
- Standardization
- Dropping categorical columns



ALGORITHMS

01

K NEAREST NEIGHBORS

classifies data points based on the majority class among their k nearest neighbors

02

RANDOM FOREST

uses multiple decision trees to enhance predictive accuracy and mitigate overfitting

03

LINEAR REGRESSION

modeling relationships between quantitative variables

04

CLUSTERING

segregates datasets into distinct clusters by iteratively assigning data points to the closest cluster centroid

RESEARCH QUESTIONS

01 Can we predict the win percentage for each team for a particular season?

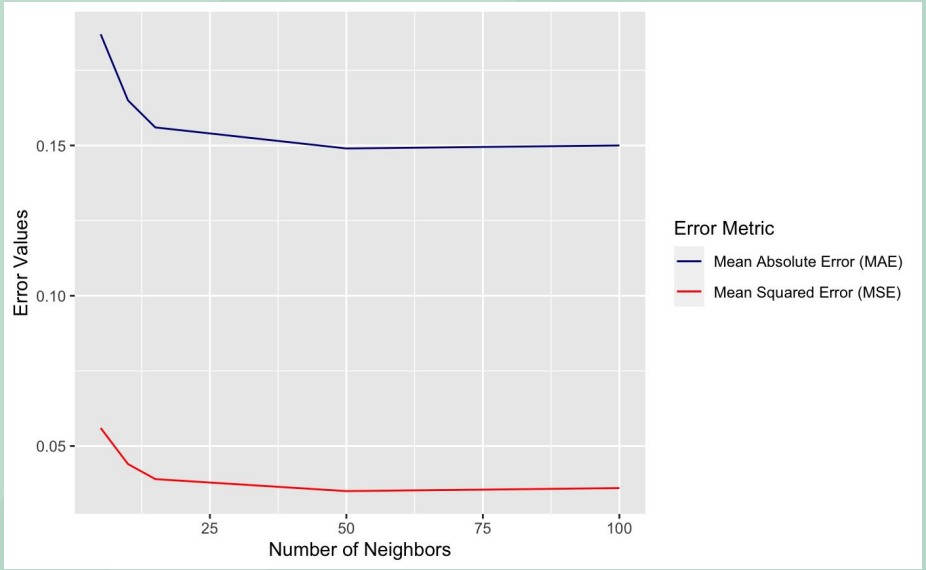
02 Can we predict whether or not a team will have a winning record given their statistics?

03 Can we predict a team's net point gains a season, given their statistics for that season?

04 Can we group teams by their playing style, such as defensive or offensive, over the 20 year span?

Can we predict the win percentage for each team for a particular season?

- K-Nearest Neighbors
- 50 nearest neighbors for defense
- 100 nearest neighbors for offensive
- Mean Absolute Error: 0.149 Defensive & Offensive



Can we predict whether or not a team will have a winning record given their statistics?

- Winning Record (1) or Losing Record (0)
- 10 attributes, 32 data points, 10 decision trees
- Accuracy: 68.5%
- Precision: 53.6%
- Recall: 80.7%

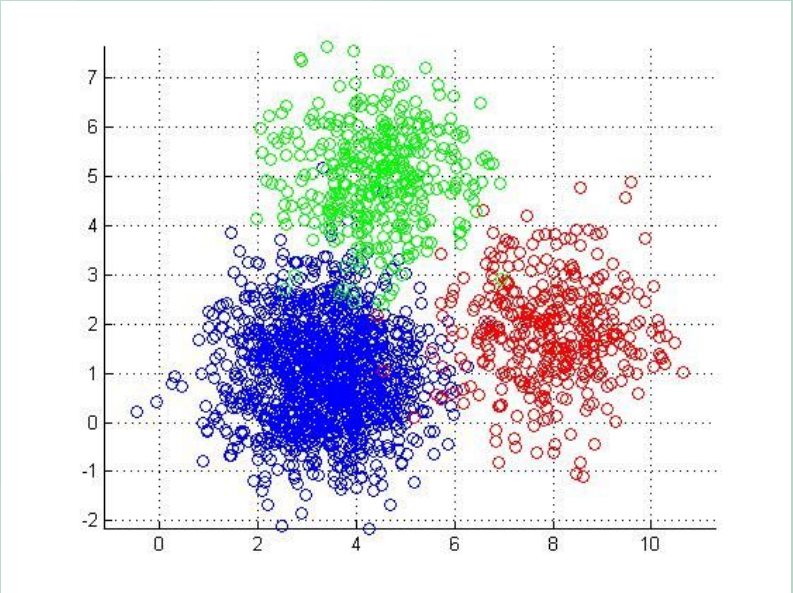
	Predicted Losing Record	Predicted Winning Record
Actual Losing Record	297	71
Actual Winning Record	141	163

Can we predict a team's net point gains a season, given their statistics for that season?

- Linear Regression
- Combines Offensive and Defensive Statistics
- Points gained - points opponents gained
- Measured in R squared

Can we group teams by their playing style, such as defensive or offensive, over the 20 year span?

- K-Means Clustering
- 10 clusters
- Standardized data
- Third cluster: “Worst” teams
- Fourth cluster: “Best” teams
- Best teams are strong on offense and defensive; worst teams are bad on offense and defensive



RESULTS

WIN PERCENTAGE
Moderate Accuracy

WINNING RECORD
Reasonable level of precision

NET POINT GAMES
Captured variance

DEFENSIVE/OFFENSIVE
Revealed distinct groupings

CONCLUSIONS

- Statistical and Machine Learning techniques
- Power of data-driven approaches in sports analytics

ANY QUESTIONS?