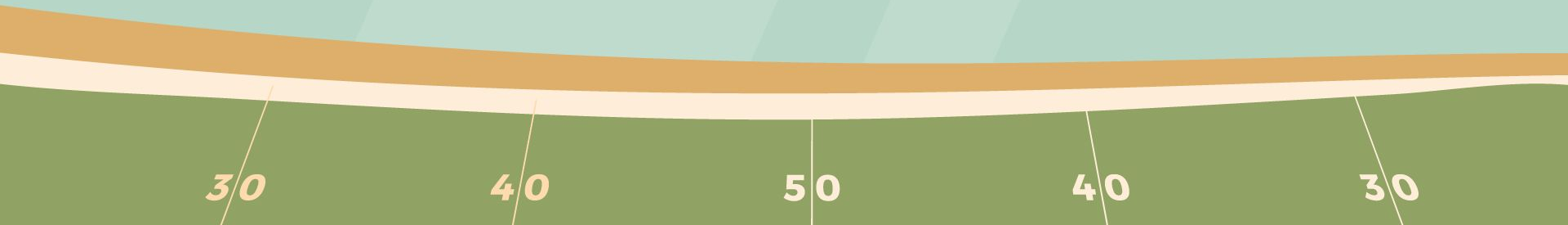
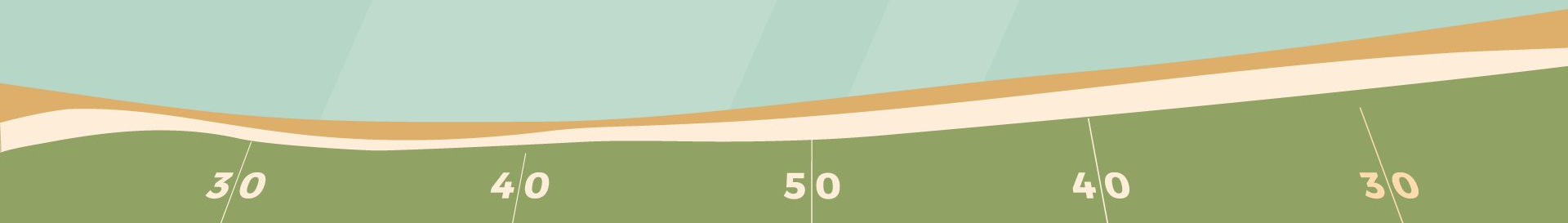
DS 340 Final Project: Beating the Books

Easwer Raman and Victor Verma



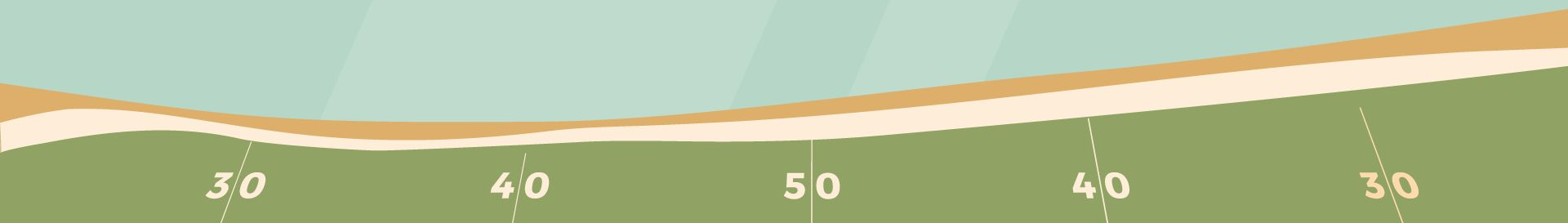
Problem Statement

- Predict the NFL game statistics of quarterbacks, running backs, wide receivers, and tight ends
- Possible uses in both fantasy football and to test the sportsbooks
- Need to take opponent defenses into consideration
- Evaluated by comparing predictions to the actual statistics of football players



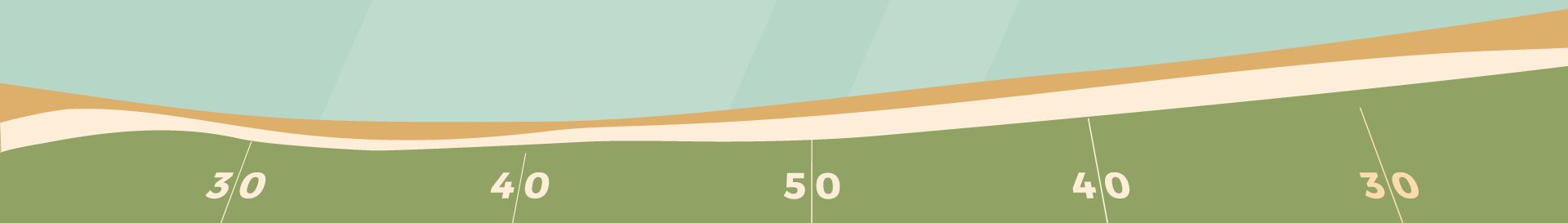
Methods

- Used a pro-football-reference scraper python package to get previous game stats in time series
- Trained both a LSTM and a Simple RNN on the time series data
- Adjusted predicted stats based on proportional opponent defensive rankings



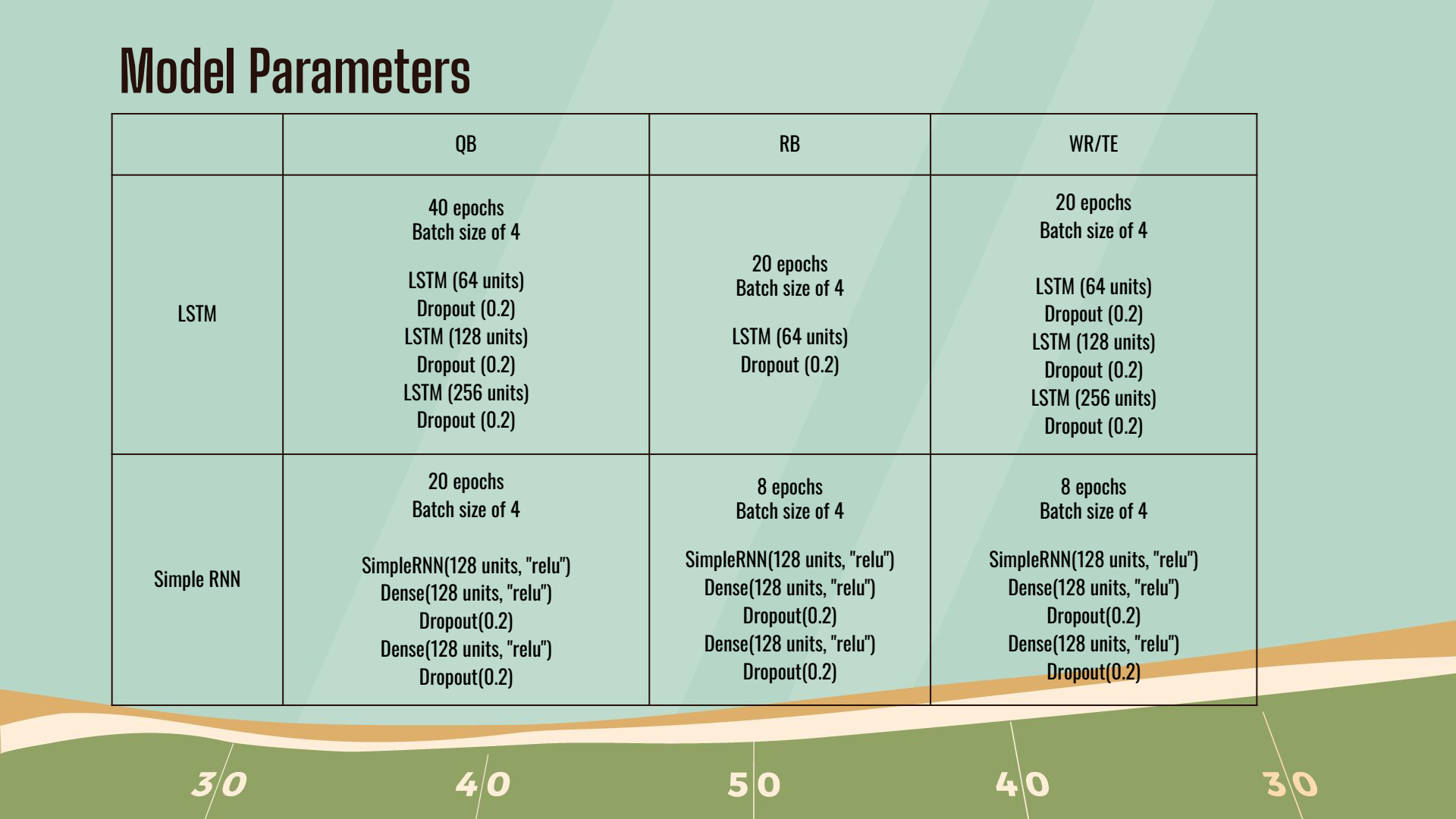
Model Features

- QB features: cmp, att, pass yards, pass tds, int, rating, sacks, rush att, rush yards, and rush tds
- RB features: rush attempts, rush yards, rush tds, targets, receiving yards, and snap percentage
- WR/TE features: targets, receptions, receiving yards, receiving touchdowns, and snap percentage



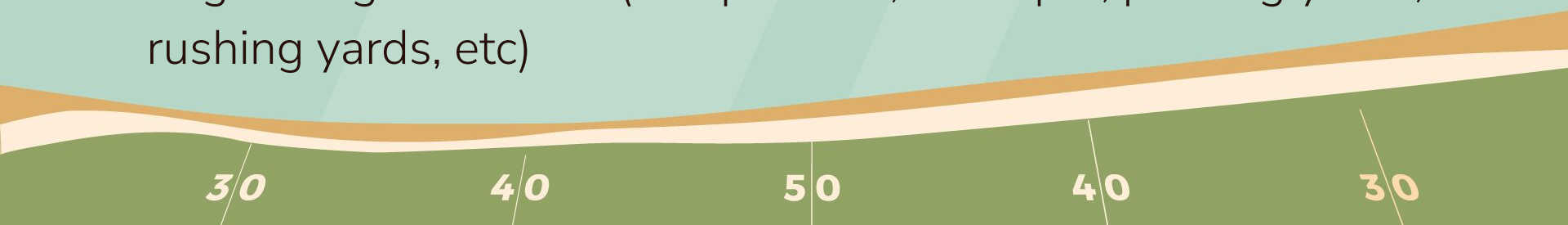
Model Parameters

	QB	RB	WR/TE
LSTM	40 epochs Batch size of 4 LSTM (64 units) Dropout (0.2) LSTM (128 units) Dropout (0.2) LSTM (256 units) Dropout (0.2)	20 epochs Batch size of 4 LSTM (64 units) Dropout (0.2)	20 epochs Batch size of 4 LSTM (64 units) Dropout (0.2) LSTM (128 units) Dropout (0.2) LSTM (256 units) Dropout (0.2)
Simple RNN	20 epochs Batch size of 4 SimpleRNN(128 units, "relu") Dense(128 units, "relu") Dropout(0.2) Dense(128 units, "relu") Dropout(0.2)	8 epochs Batch size of 4 SimpleRNN(128 units, "relu") Dense(128 units, "relu") Dropout(0.2) Dense(128 units, "relu") Dropout(0.2)	8 epochs Batch size of 4 SimpleRNN(128 units, "relu") Dense(128 units, "relu") Dropout(0.2) Dense(128 units, "relu") Dropout(0.2)



Challenges

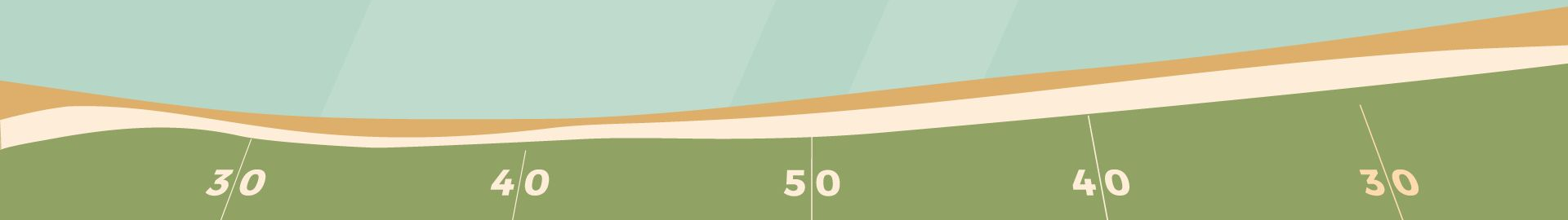
- Scraping training data from the internet
- Hard to balance minimizing loss and overfitting due to the small amount of source data (weekly statistics throughout the season)
- Taking opponent defensive stats into account
- Statistics with a smaller range of values (touchdowns, interceptions, etc) were much more unreliable than those with a larger range of values (completions, attempts, passing yards, rushing yards, etc)



Results

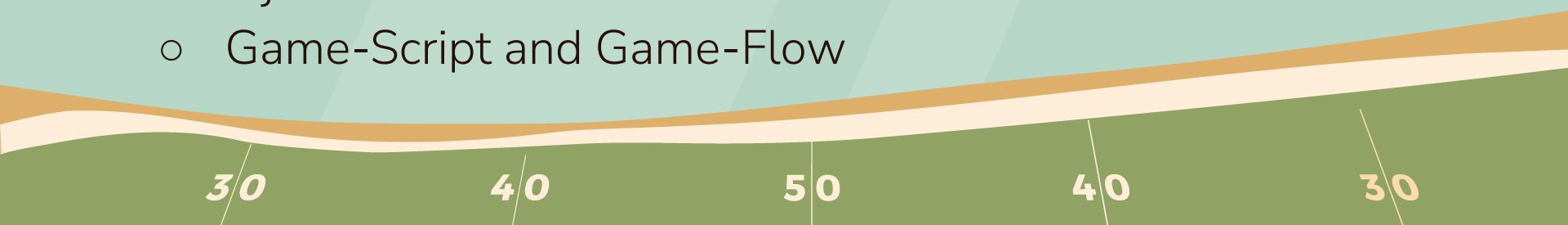
- LSTM usually works better than the Simple RNN
- Implementation of defensive stats helps with TD and Int variation

	Completions	Attempts	Pass Yards	Pass TD	Interceptions	Rating	Sacks	Rush Attempts	Rush Yards	Rush TD
LSTM	22.4	33.3	218.4	1.32	0.47	89.4	1.6	4.2	19.0	0.0
Simple RNN	22.9	34.6	239.2	1.97	0.1	89.4	1.6	3	8.7	0.0
Actual	21	33	210	1	1	79.1	3	4	26	0



Conclusions

- The LSTM performed better than the RNN
- Statistics with a greater range were easier to predict
- We were able to beat the books with around 66.7% accuracy (32/48), without injuries, we had 74.4% accuracy (32/43)
- There are many important factors that impact NFL games that are not possible to model
  - Injuries
  - Game-Script and Game-Flow



Fun Fact

- We thought the algorithm would be wrong about the passing yards and passing touchdowns for the Cowboys Eagles game