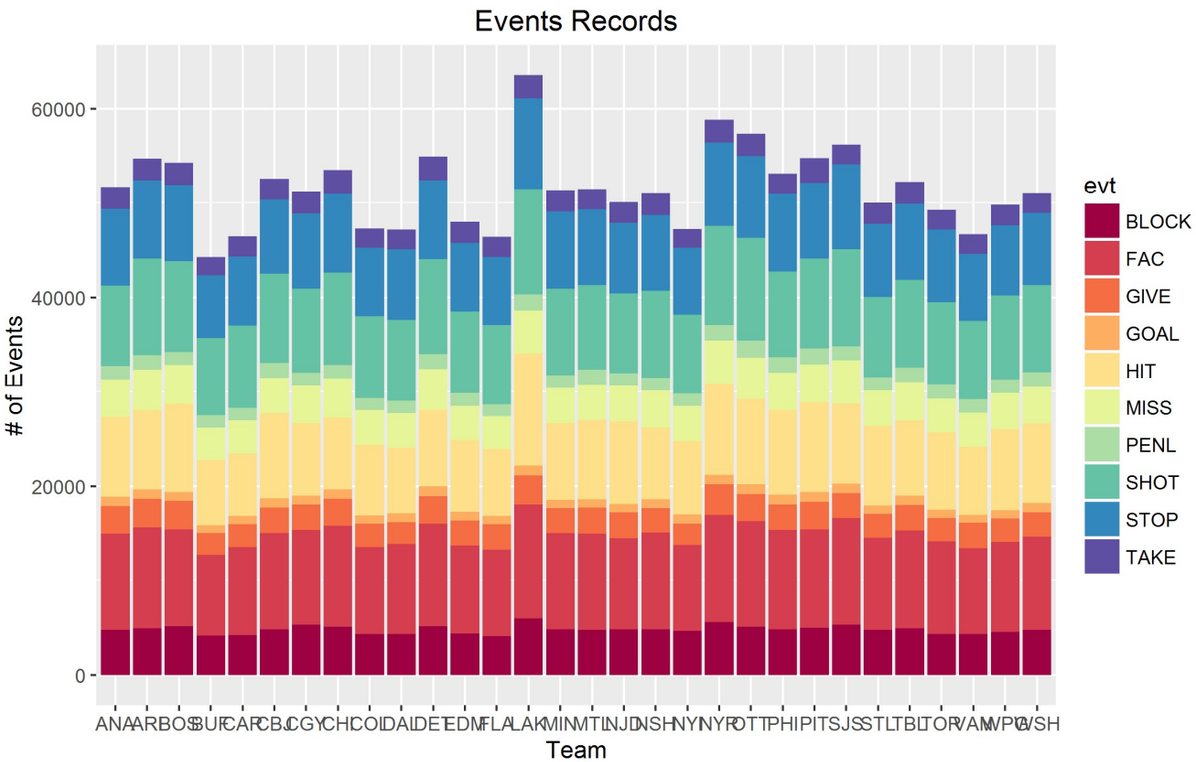
# Project NHL

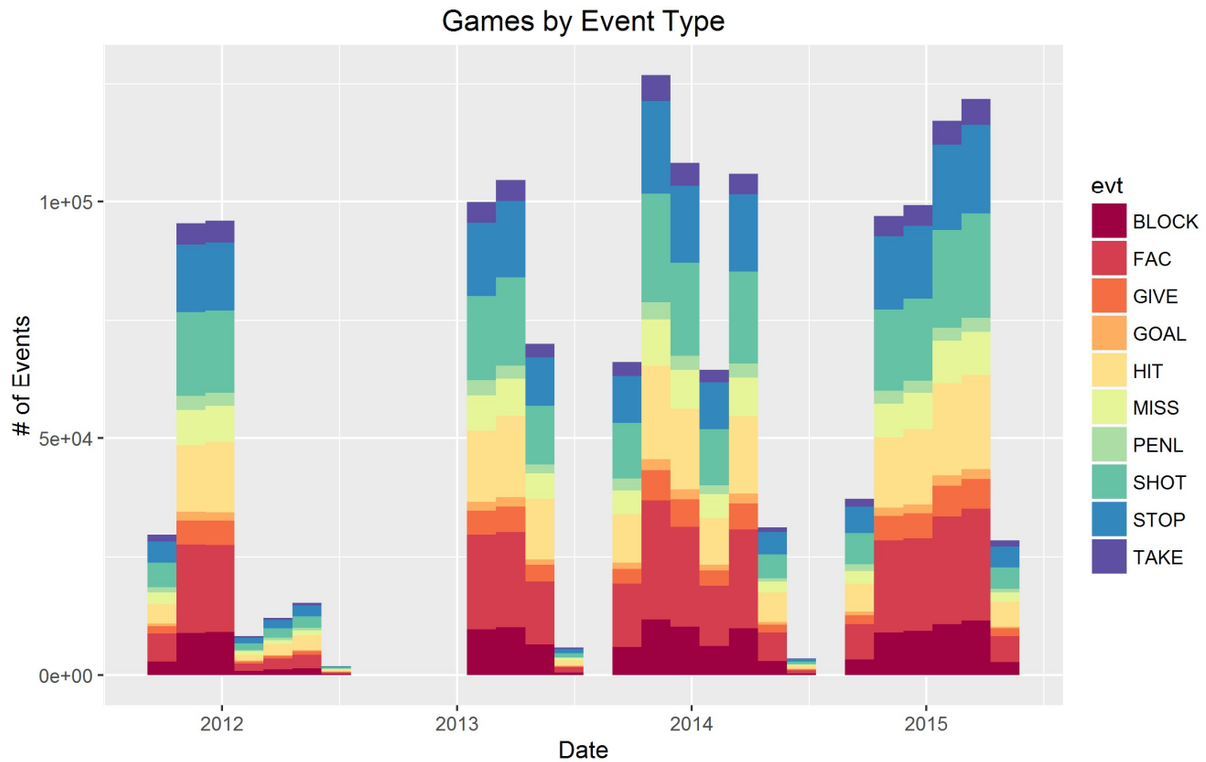
By Steven Ginzberg

# Scraping NHL.com

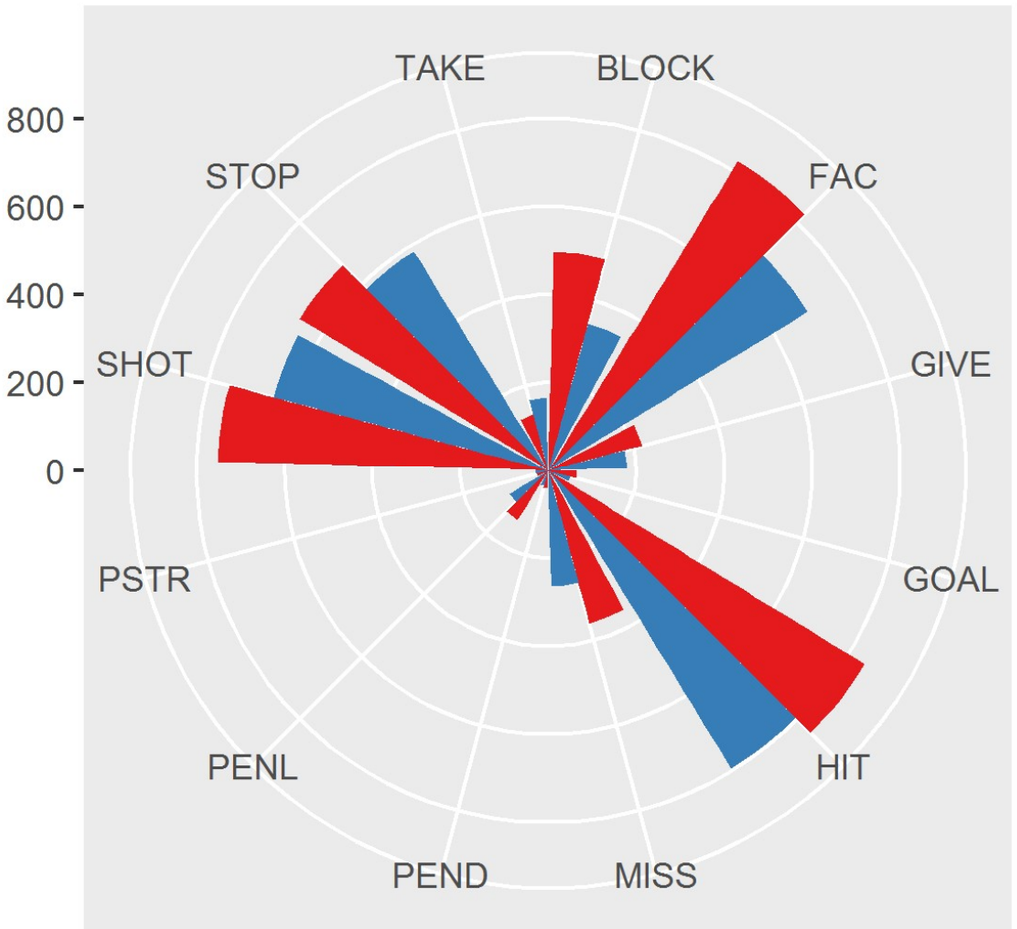
- Scraped from 2012-13 through 2015-16
  - Gap in games during 2012-13 for the NHL lockout
- Received all events for 4,308 out of a possible 9600 games
- Event types include Goals, Penalties, Face-offs, Shots, Blocks, etc.
- 1.765 million records and 56 fields
- Some data is missing relating to game header information
  - Approx. 170,000 records, or 10%

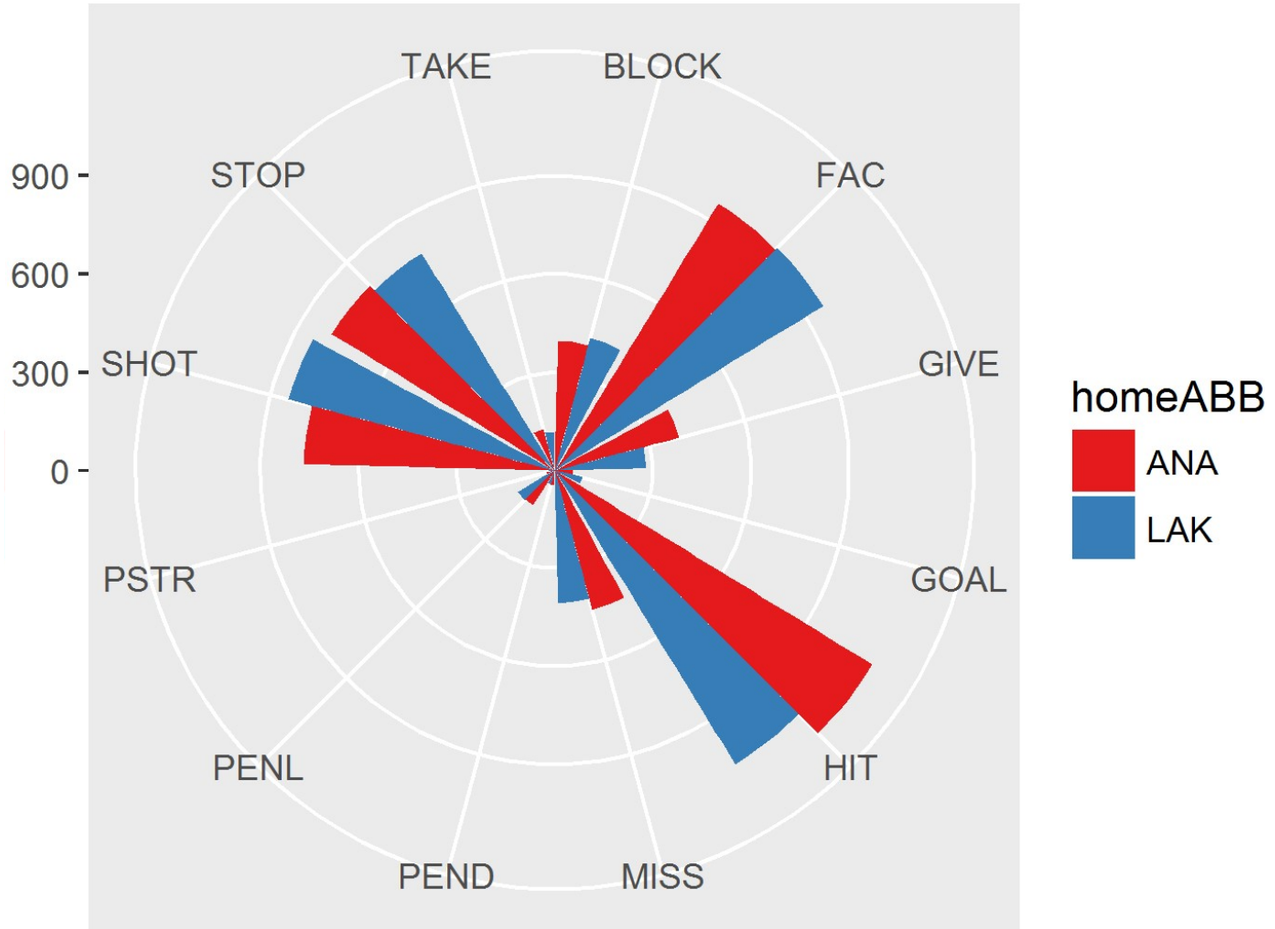
# Database Visually



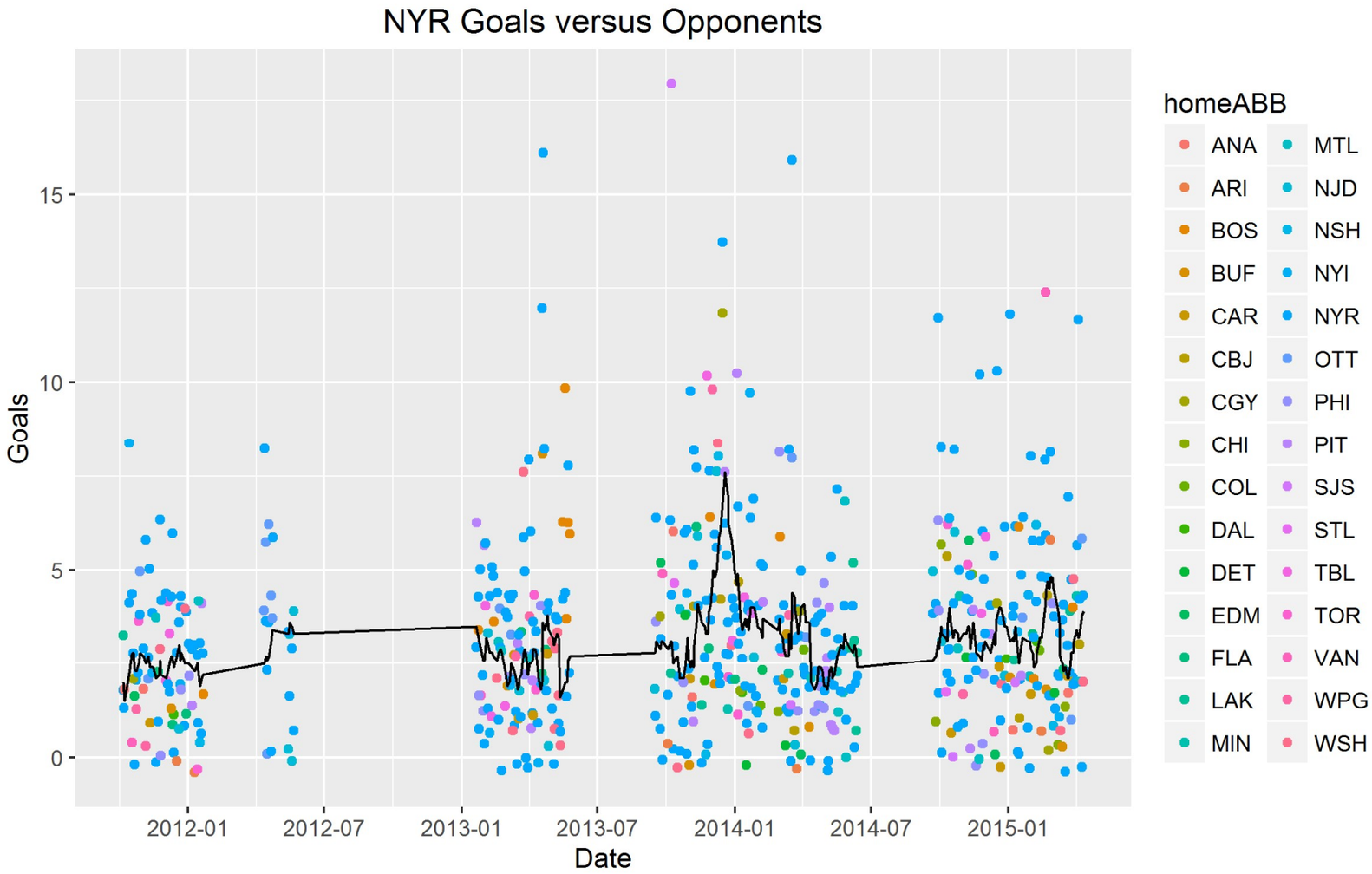



# Database Visually



# Visualizing Moving Averages



# Machine Learning Algorithms

- The objective is to learn from a series of events and try to predict the likelihood of a goal.
- Using event database required setting up a separate column for goal and removing the goal events.
- In the best case, I was expecting change in odds to be minimal
- Naïve-Bayes Model:
  - Uses all available evidence
  - Works well with categorical and text-based data
  - Result was it predicted NO GOALS
    - the likelihood of a goal in any case is very small so it always went to zero.
- Logistic Model:
  - Attempts to get probabilities failed – needed a huge amount of memory

# Going Forward

- Fix current issues:
  - I need to scrape for the missing data and set up an application to scrape during the season
  - Switch to an Linear or Logistic model that can handle both categorical and continuous variables and provide odds of success
- Future issues:
  - Break down text fields to include player by player data – great for fantasy sports!
  - Additional visuals during the course of the season