Project Overview

- Multi-year, entrepreneurial (personally derived) effort in football data analytics.
- Initially focused to leverage data and to create a proprietary performance metric to communicate performance in a past event (objective and subjective).
- Those past events, evaluated thru the performance metric, were successfully integrated as the foundational element of a predictive model for game outcomes.
- Organic collaborations and partnerships. This has become a validated proof of concept and representation of personal grit.

“The toughest tests are the ones we give ourselves”
-found in the Cal Poly ROTC hallway

Personal Overview

- Attended and played football here at Cal Poly. Coached for 10 years at Cal Poly and Army. This created a unique domain specific knowledge opportunity.
- Inclination towards technical analysis, mathematics and computer coding.
- Combination of experience, knowledge and passion for applied science and football. (do something with proper theory and fundamentals)
- The analytics “team” - math person, coder and football person. Staff meetings should be easy to schedule….

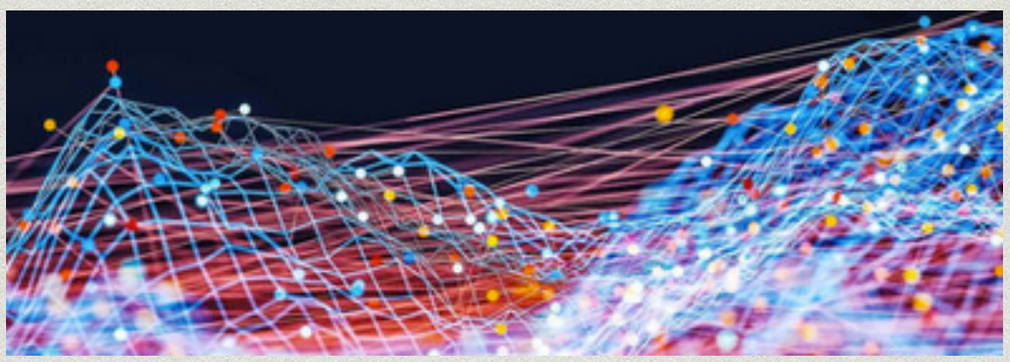
Analytics Overview

- Broad concept: The discovery, interpretation and communication of patterns in data.
- Analytics Maturity -

1. Descriptive Analytics : Hindsight
2. Diagnostic Analytics : Oversight- What is happening and why
3. Predictive Analytics : Foresight- What will happen
4. Prescriptive Analytics : Insight- How can we optimize what is happening?
5. Cognitive Analytics : Generate new questions and hypotheses

Football Data

- Player tracking data collected in real-time
- Play data (~150/game)
- Drive data (~25/game)
- Box scores and bulk statistics (~250 games per season)
- Season statistics (100 seasons)
- Career statistics
- Franchise statistics
- League statistics



Data

- Capture all the game drive data in proper sequence.
- Evaluate the drive data within context of starting field position and game situation. A contextual evaluation
- Captured data by web scraping with javascript, acquiring historical databases, processing play-by-play data to create drive data and from public facing websites.
- Size of database: over 120,000 lines of drive data and ~5000 games.
- Data management is critical - those 25 lines of data must be correct (quality input maximizes your a chance for quality output)

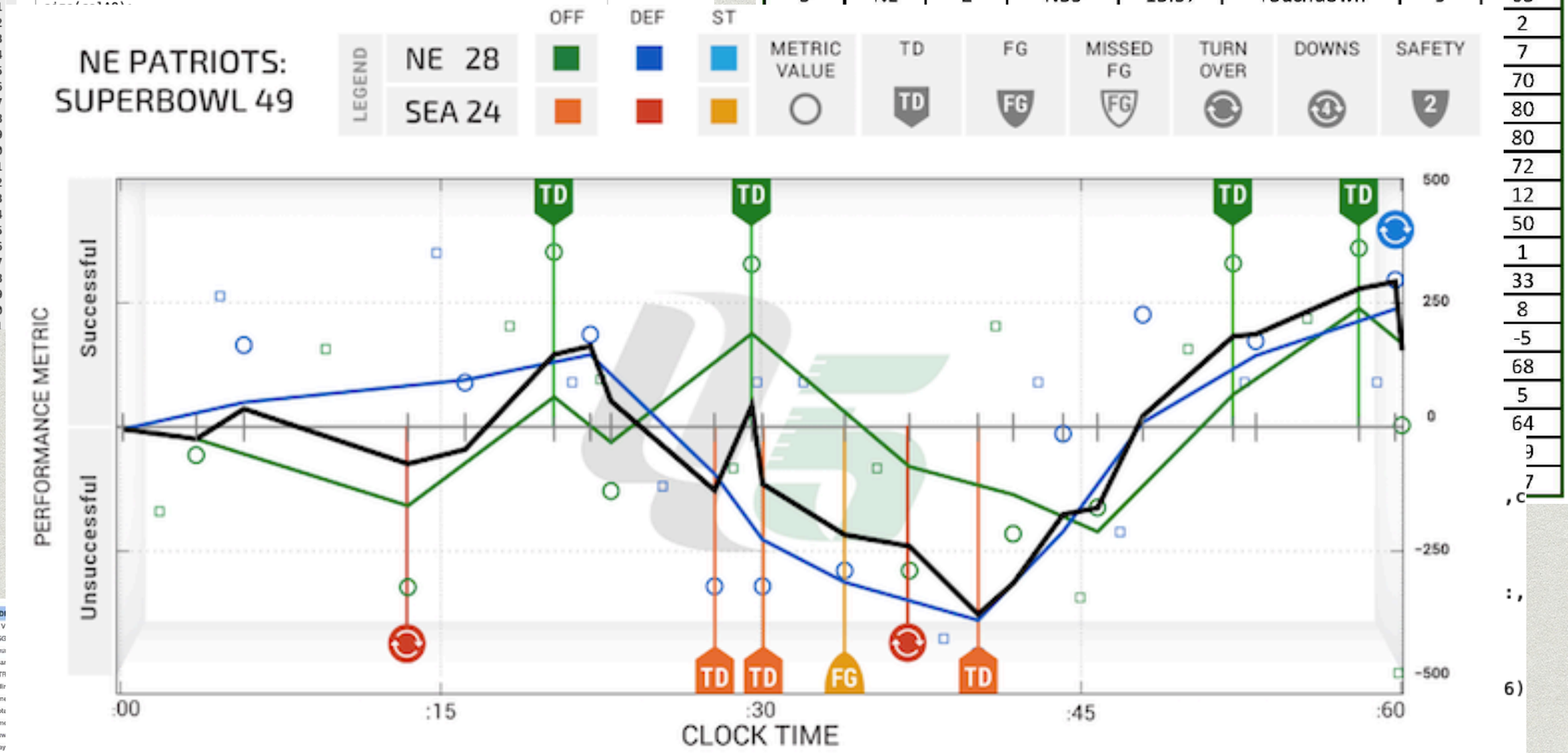
Coding the Performance Metric and Data Visualization

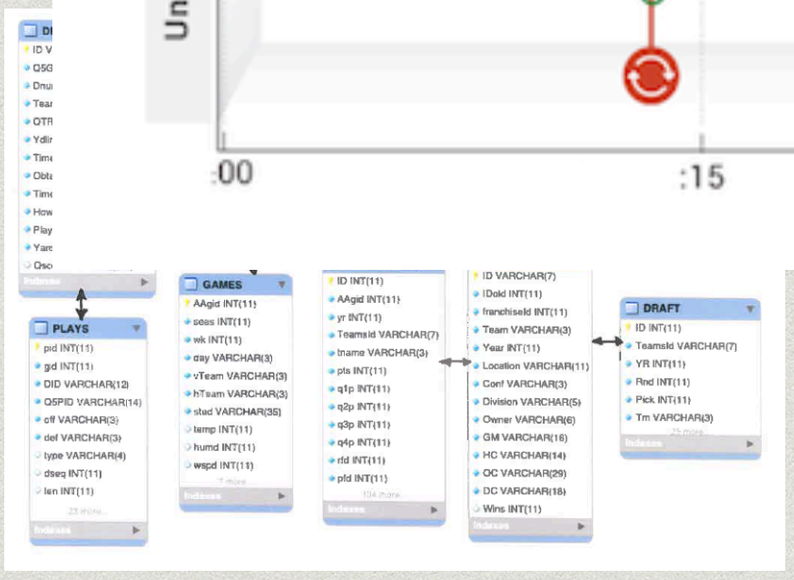
- The data was processed thru the performance metric. Placed the evaluation of each drive on a single success continuum.
- Created a time history canvas to visually display results.
- Step-by-step video explaining the components of the GameMap.
- Resourced feedback from coaching contacts in the profession.

NFLRegWeekSummaryPICSFILE.m | modelrun.m | objective22.m | objective.m

```
[~,Ao]=sortrows(modeluseddd(weeks(week,1):weeks(week,2),12));
%if week==4 Ao, pause,  end
weekSOLN1ago(Ao,1)=weekSOLN;

colAL=weekSOLN;

AMthruAP=[weekSOLN1ago weekSOLN2ago weekSOLN3ago weekSOLN4ago];

colAQ=weekSOLN-weekSOLN1ago;
```

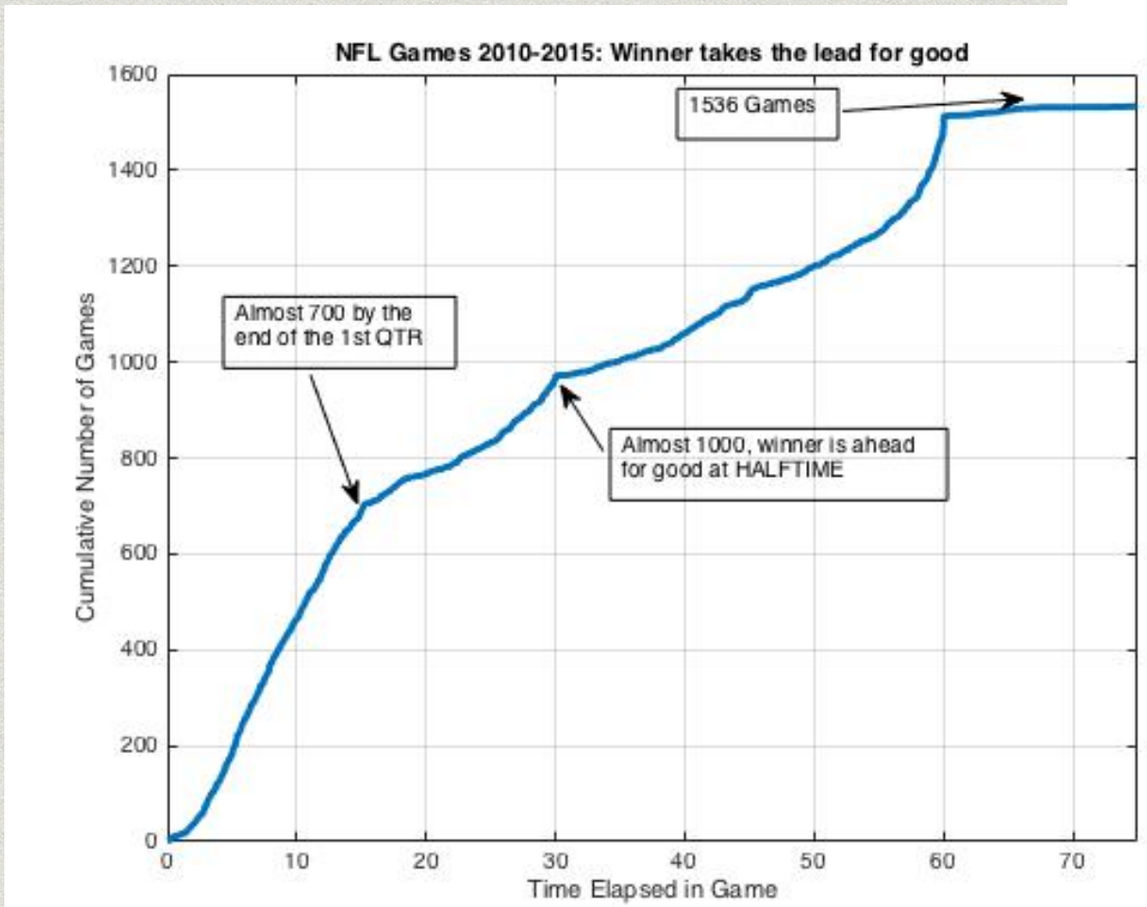
Drive	Team	Qtr	Yard Line	Time	HowEnded	Plays	Yds
1	NE	1	N18	15:00	Punt	6	17
2	SEA	1	S16	11:34	Punt	3	8
3	NE	1	N32	9:20	Interception	13	58
4	SEA	1	S14	1:39	Punt	5	12
5	NE	2	N35	13:57	Touchdown	9	65
							2
							7
							70
							80
							80
							72
							12
							50
							1
							33
							8
							-5
							68
							5
							64
							9
							7

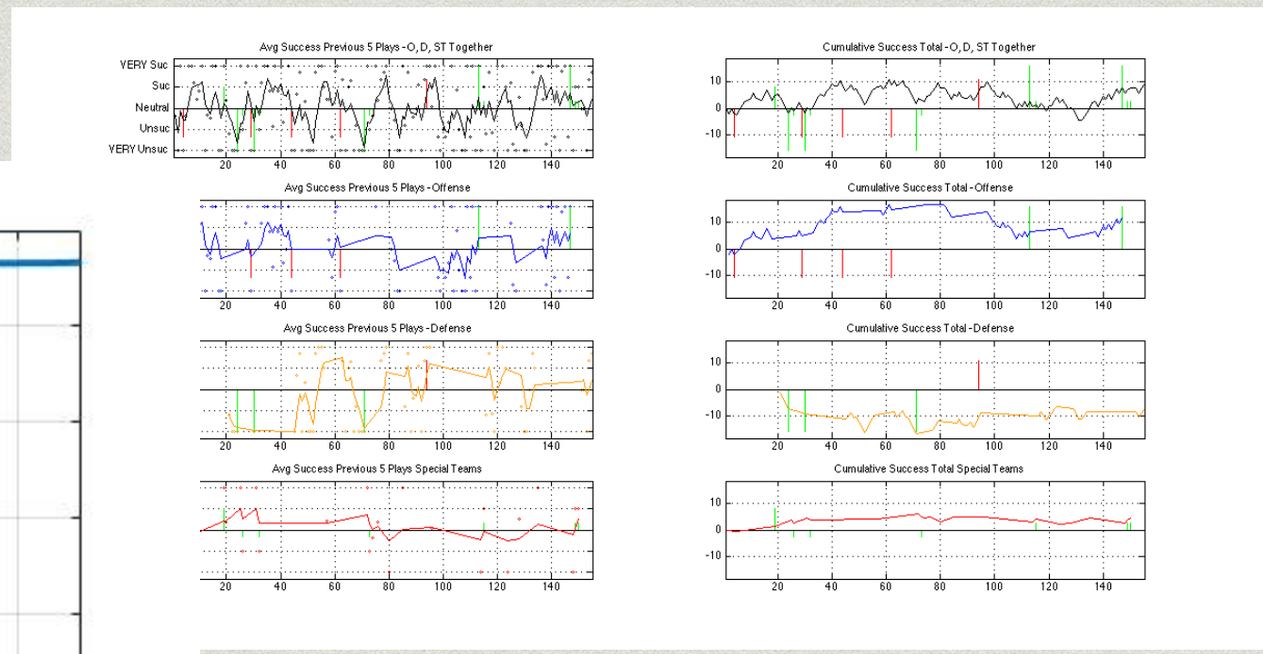


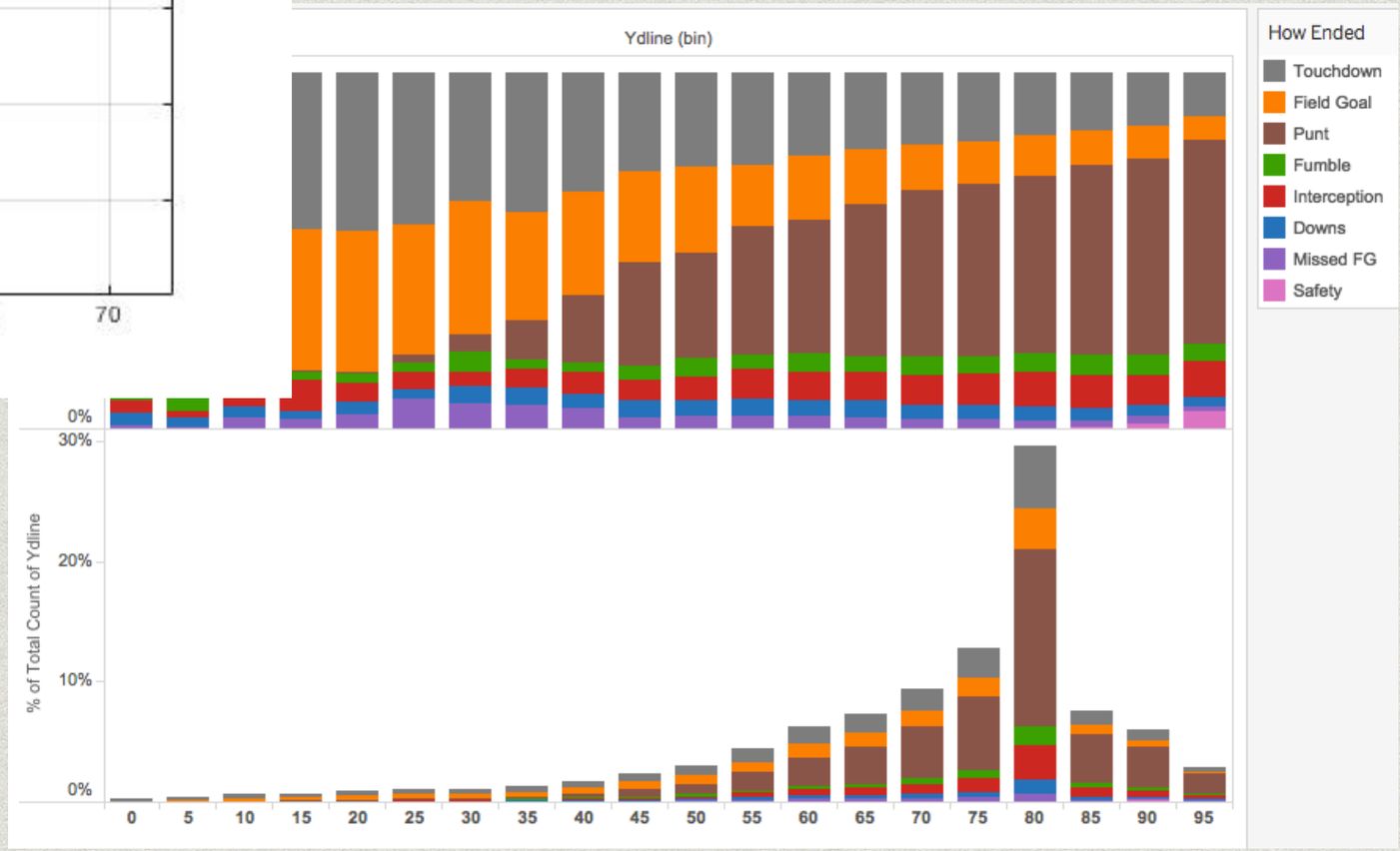


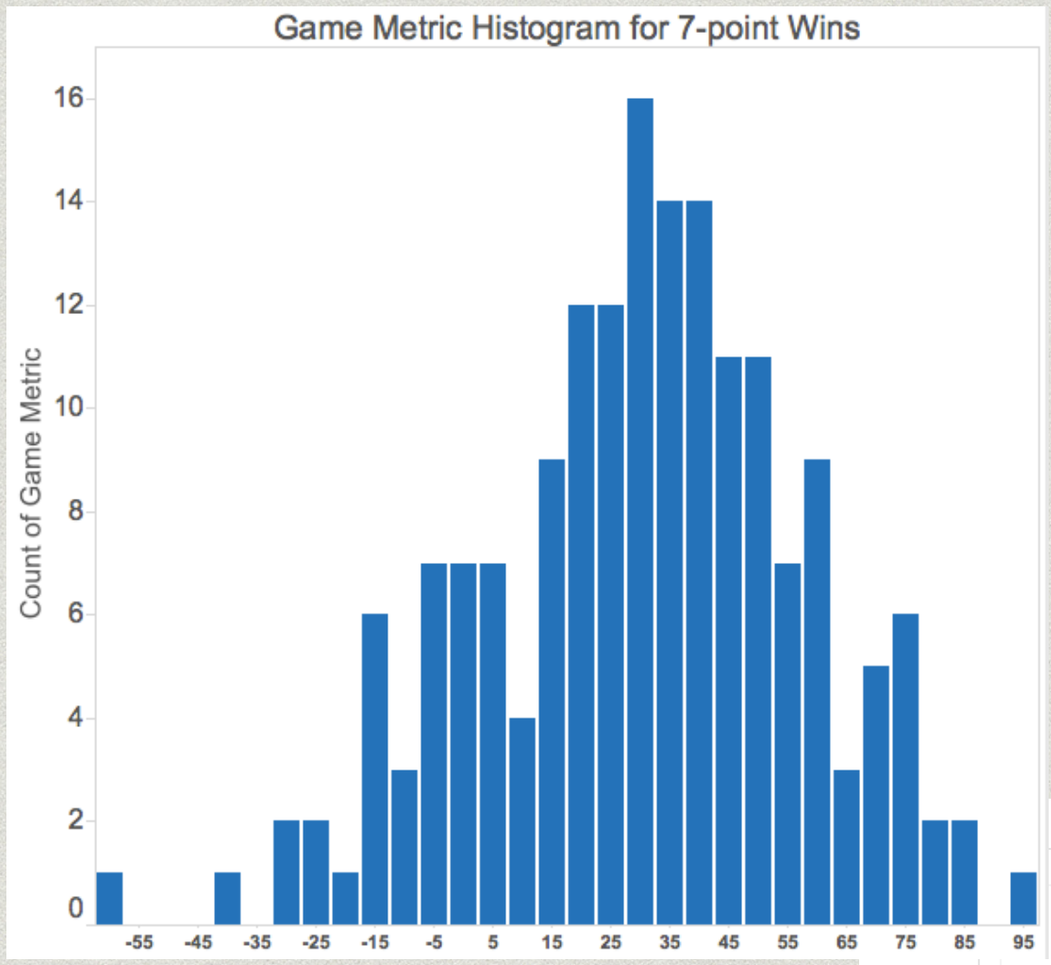

```
teamtestlose=['teamsnames' teamLOSE]';
teamLOSEkeep=cell2dataset(teamtestlose);
%%%right now A has 14 cols
HELLOwin=horzcat(teamWINkeep,SucWIN);

VALdataTEAM1=VALdataTEAM2;
VALdatawin=VALdatalose;
SucLOSE=horzcat(mat2dataset(VALKeep),mat2dataset(VALdatawin),
```

Data Visualizations



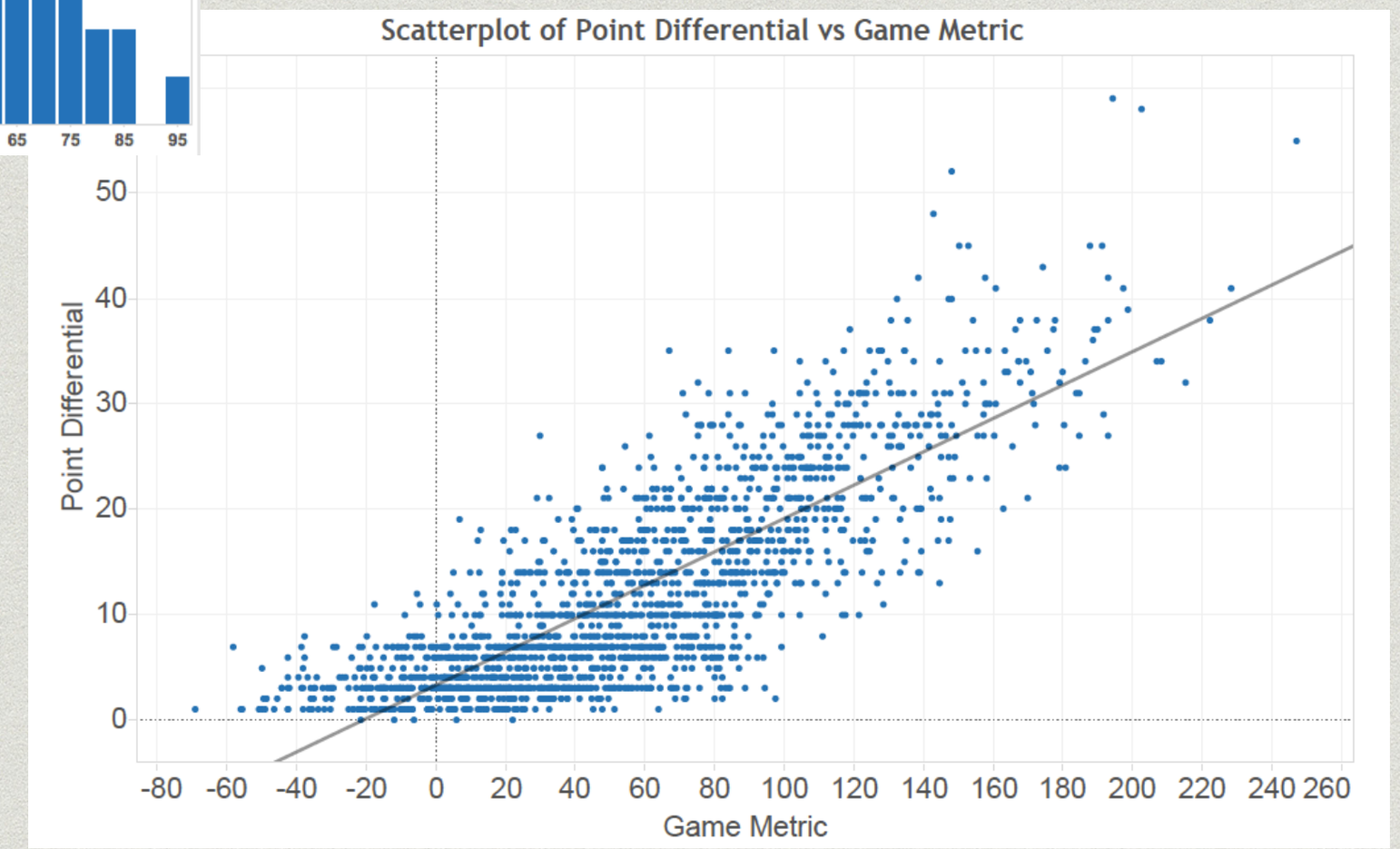






Points or Performance?

Can Past Performance Predict The Future?



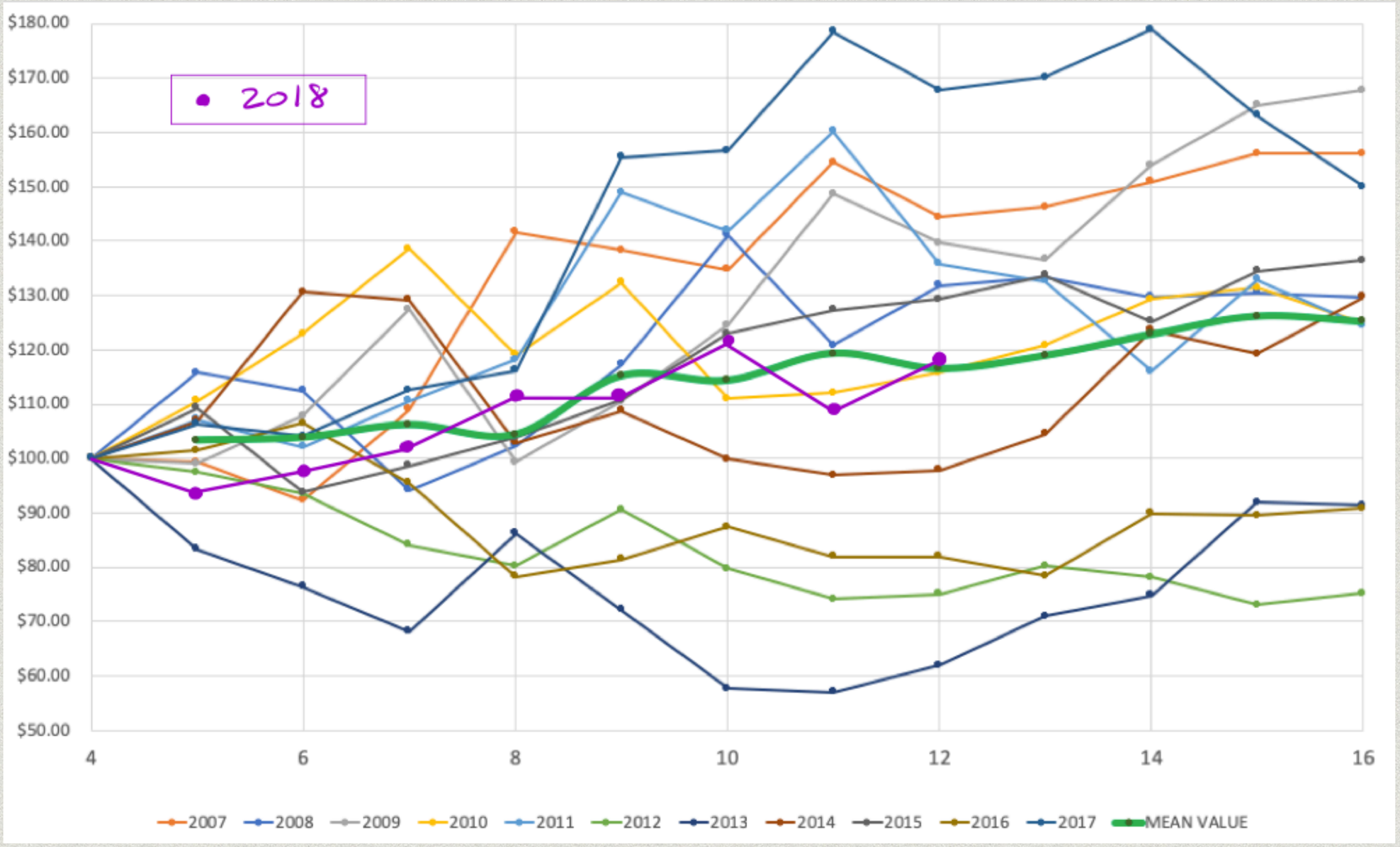
Diagnostic Analytics

- Overall team strength (rankings).
- Team components.
- Performance within situational context at the drive level.
- What are teams doing that “win” - the anatomy of winning football

Building a Predictive Model

- Performance evaluation attached to teams from each game - replaced scoreboard numbers.
- Use season-to-date and created novel solution over past 4 weeks to account for strength of schedule.
- Solution using a nonlinear minimization algorithm with constraints.
- Domain specific knowledge to create and tune model parameters.
- Operates on a subset of games where value is identified from an existing marketplace.

Results of Backtesting Real-Time Validation



Conclusions

- The performance metric is validated. We should trust the descriptions it produces. The descriptive analytics produced have value and should be used to evaluate real-time analysis.
- This data-driven predictive model with no human intervention contains edge in football markets.
- Past performance has been successfully used to predictive future outcomes in a force-on-force setting.
- This modeling process is scalable.

Improvements

- Further R&D is warranted for this successful baseline setting.
- Further automation is warranted.
- Interactive visualizations - links to video clips. User experience venture.
- Win probabilities should be investigated - both before and during game analysis.