COC Squad Fantasy Football
Report for 2023

Zachariah Irwin

2023/12/30

Contents

1 Introduction

Disclaimer: Data was scrubbed by hand from the Sleeper app. Thus, there may be some errors, but for the large part, the data set can be trusted.

This report is organized into five sections. In Section 2, team-by-team performance of actual scoring data is analyzed. In Section 3, team-by-team performance of *projected* scoring data—that is, whatever the Sleeper projected score based on a team's starting lineup prior to the MNF game—is compared to *actual* scoring data and analyzed. Following this, Section 4 shows team-by-team performance of *possible* scoring data–that is, what a team's score would be given an optimal starting lineup–compared to *actual* scoring data. Section 5 provides information about matchup point differentials, i.e., by how many points a team beat or lost to another team. Lastly, Section 6 includes a regression analysis to test whether or not there is correlation between total points for, point variance (from a team's average performance), and a team's record.

1.1 A review of box plots

Frequent use has been made of the box plot to understand season-long data. Figure 1.1 shows a typical box plot for random data.

The lowest quartile, or whisker, represented by the bottommost line, is a lower bound of the data set, excluding any outliers. The first quartile is the bottom edge of the box, and is a lower bound for 25% of the data. The second quartile is the top edge of the box, and is an upper bound for 75% of the data. The highest quartile, or whisker, represented by the topmost line, is an upper bound on 95% of the data set.

Generally speaking, the smaller the box and the shorter the whiskers, the more consistent that data set is. Conversely, larger boxes and longer whiskers indicate much higher variability. For Fantasy Football, a smaller box plot located above a league average for the measure of interest (which herein is denoted by a dashed black line), indicates better performance for that team in the measure of interest as compared to the rest of the league.
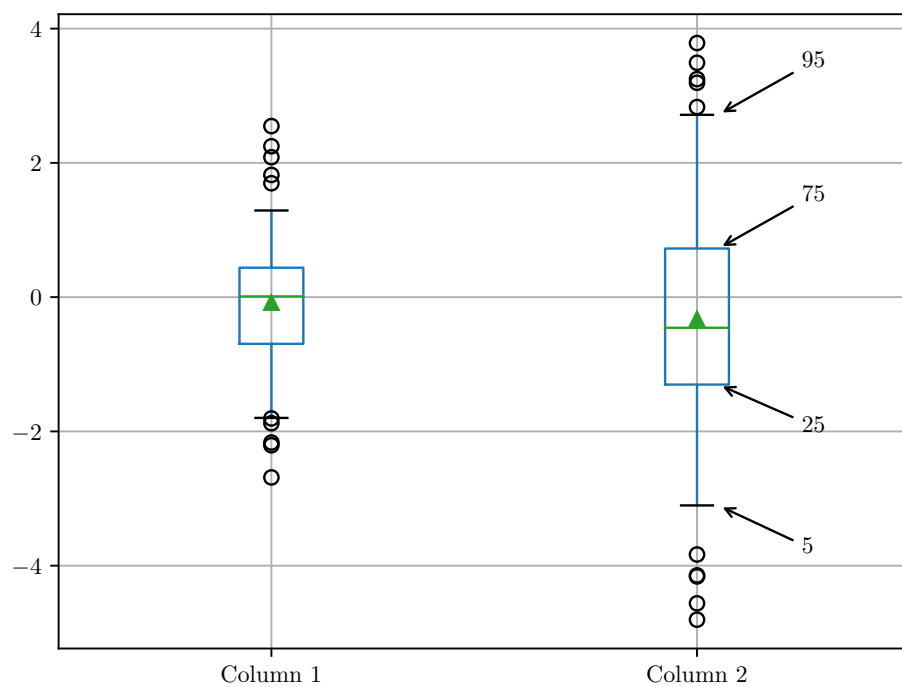
Figure 1.1: A fake data set illustrating components of a typical box plot. Green line indicates the median, and green box indicates the mean.
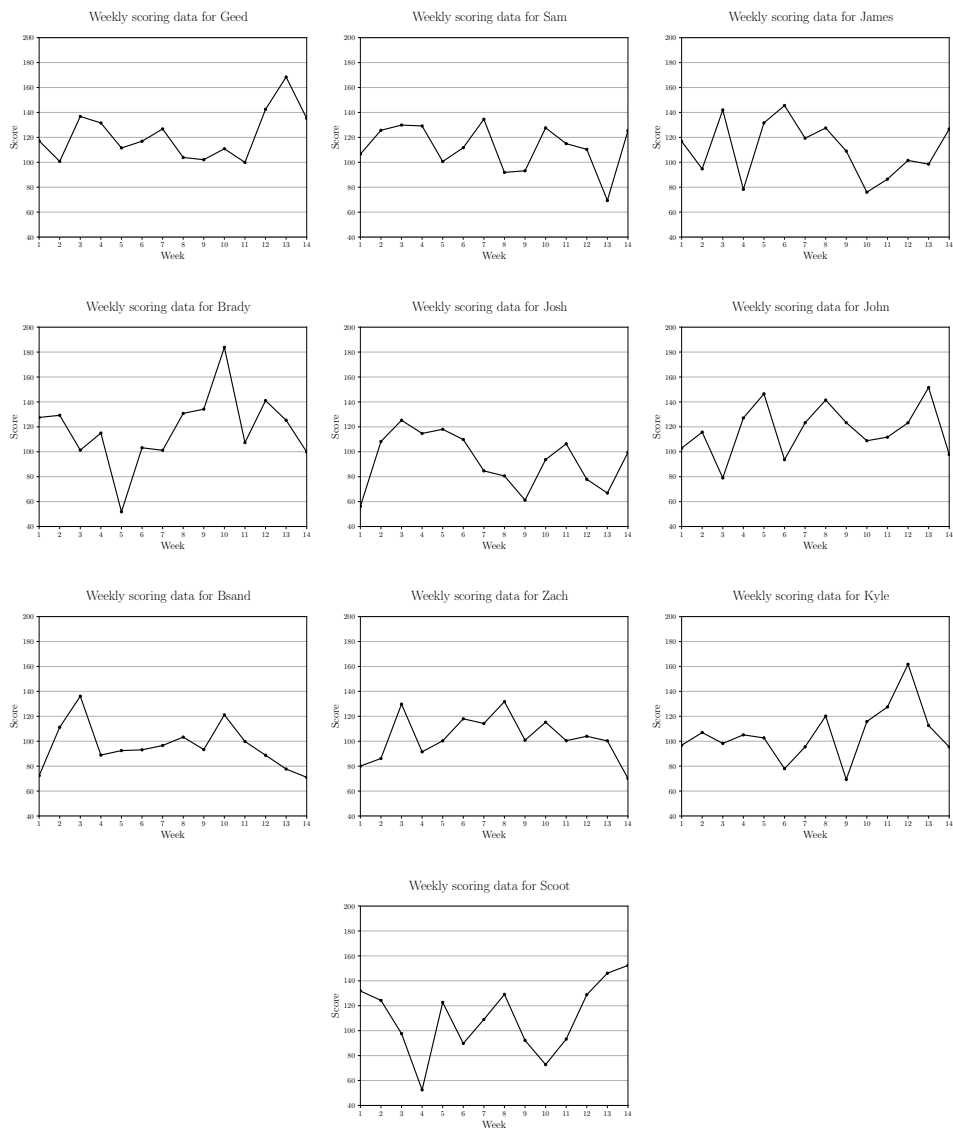
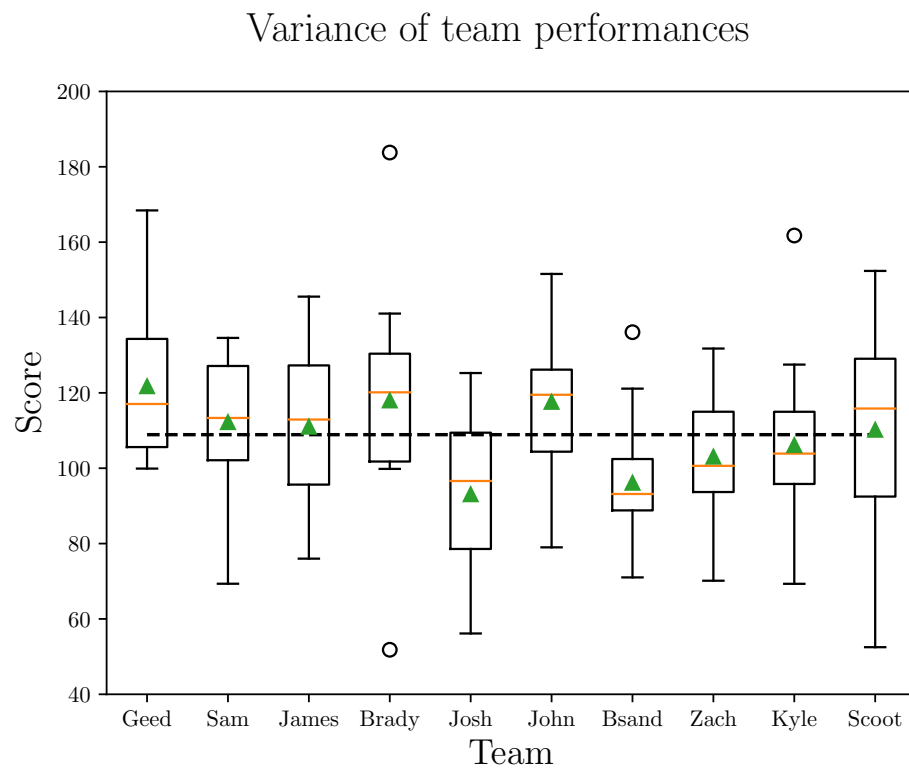# 2 Actual scores

Figure 2.1: Team scoring week-by-week.

Figure 2.2: Variance of team performances over the duration of the season. Dashed black line indicates the league average score.
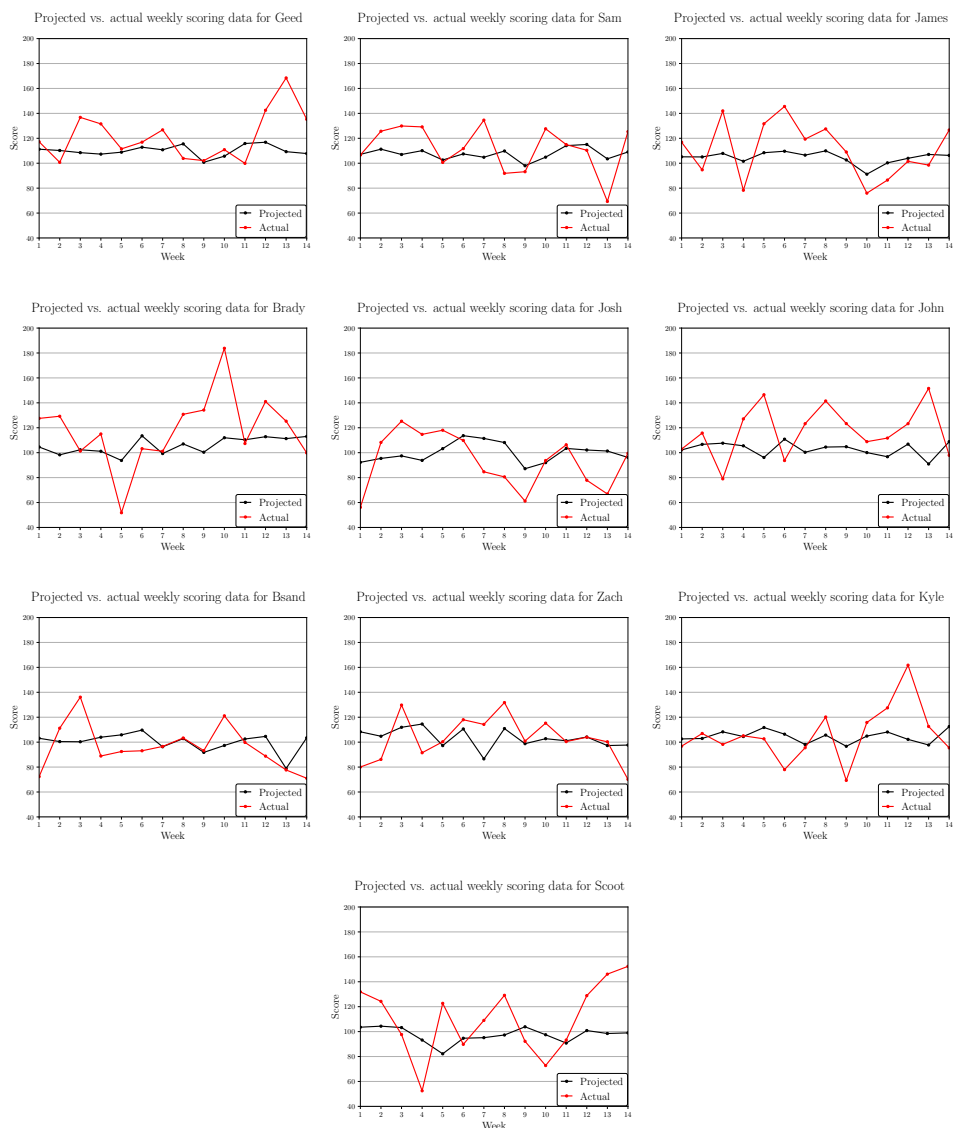
3   Projected scores

Figure 3.1: Projected team scoring week-by-week. Black lines indicate the Sleeper projection generated pre-kickoffs, and red lines indicate the actual score.

Figure 3.2: Variance of extitprojected team performances over the duration of the season. Dashed black line indicates the league average projected score.
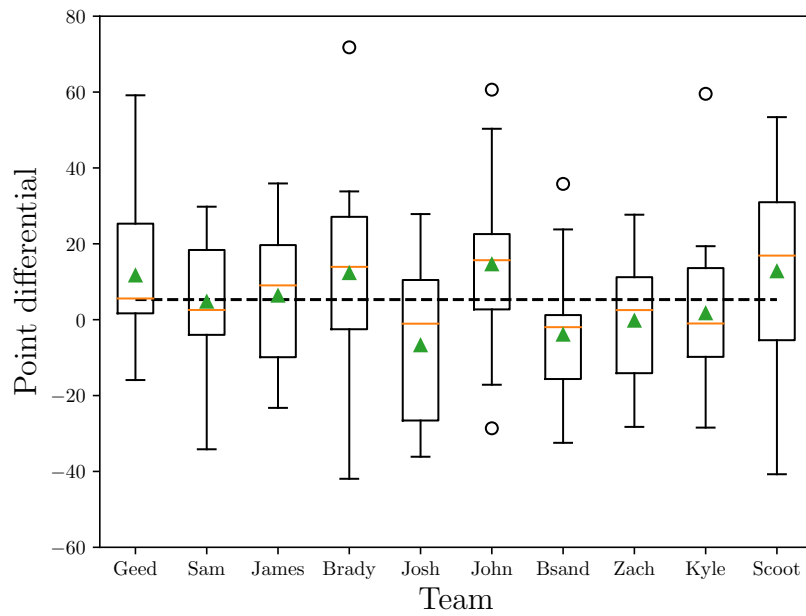
Figure 3.3: Variance of extitprojected team performances subtracted from extitactual team scores over the duration of the season. Dashed black line indicates the league average differential, which was positive over the season, meaning that on average everyone out-performed the Sleeper projection. A higher number indicates that a team out-performed projection, while a lower number indicates a team under-performed projection.

4   Possible scores
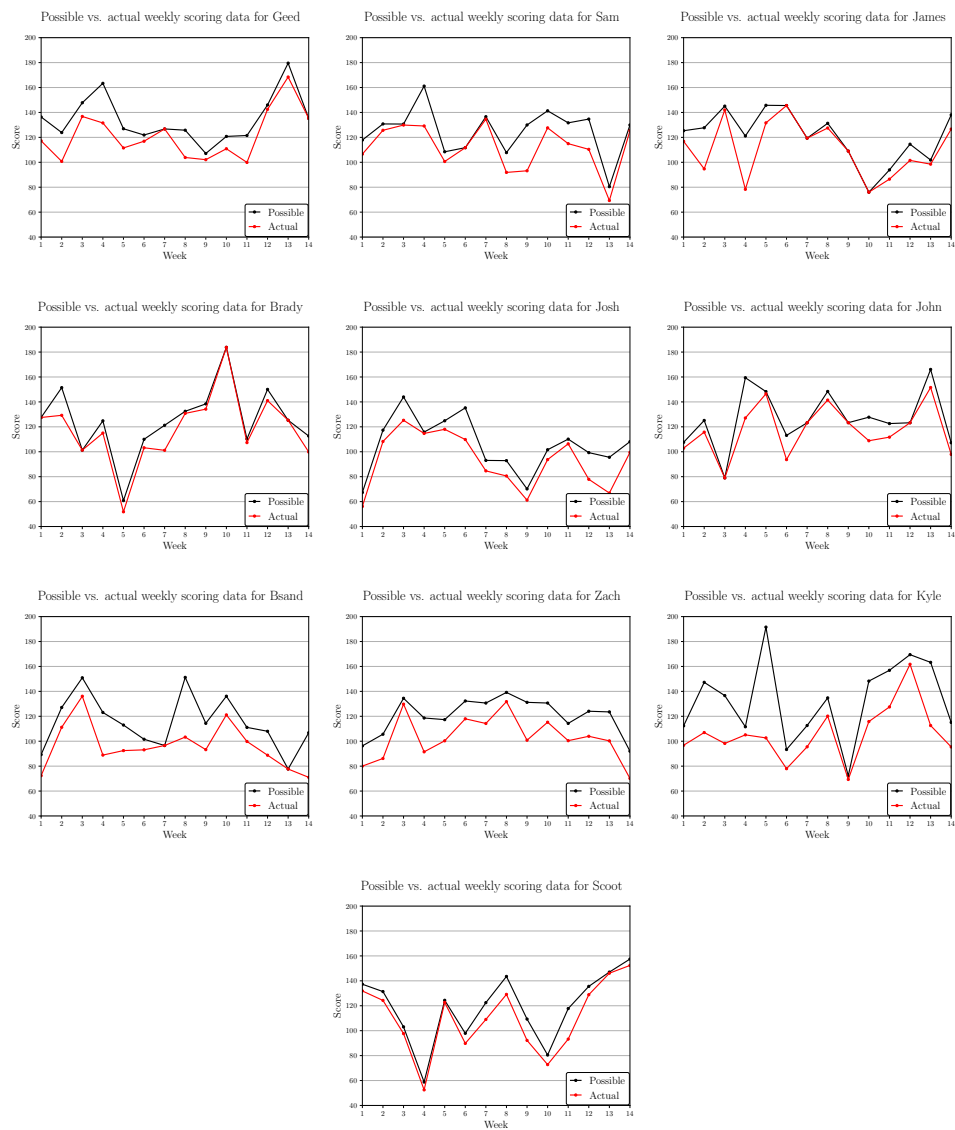
Figure 4.1: Possible team scoring week-by-week. Black lines indicate the score given an optimal starting lineup, and red lines indicate the actual score.
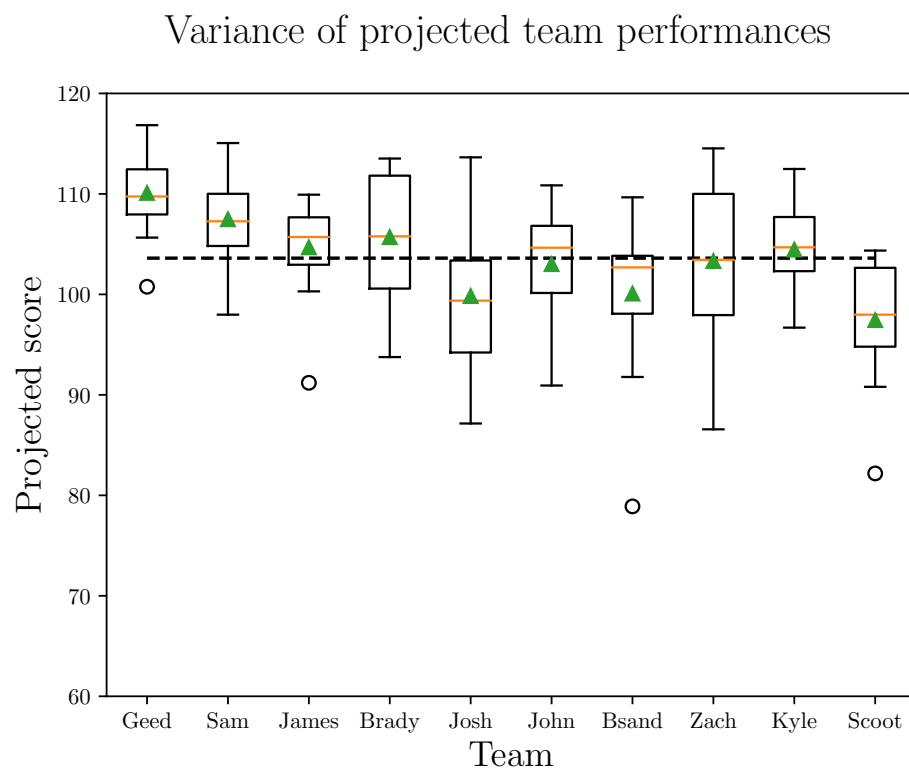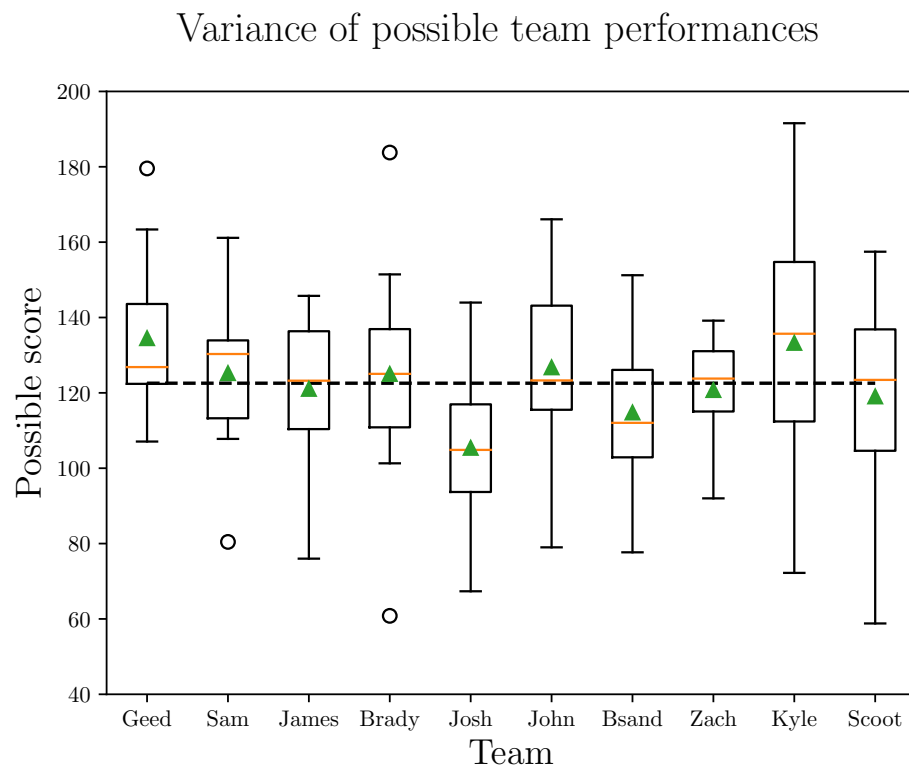
Figure 4.2: Variance of extitpossible team performances over the duration of the season. Dashed black line indicates the league average possible score.
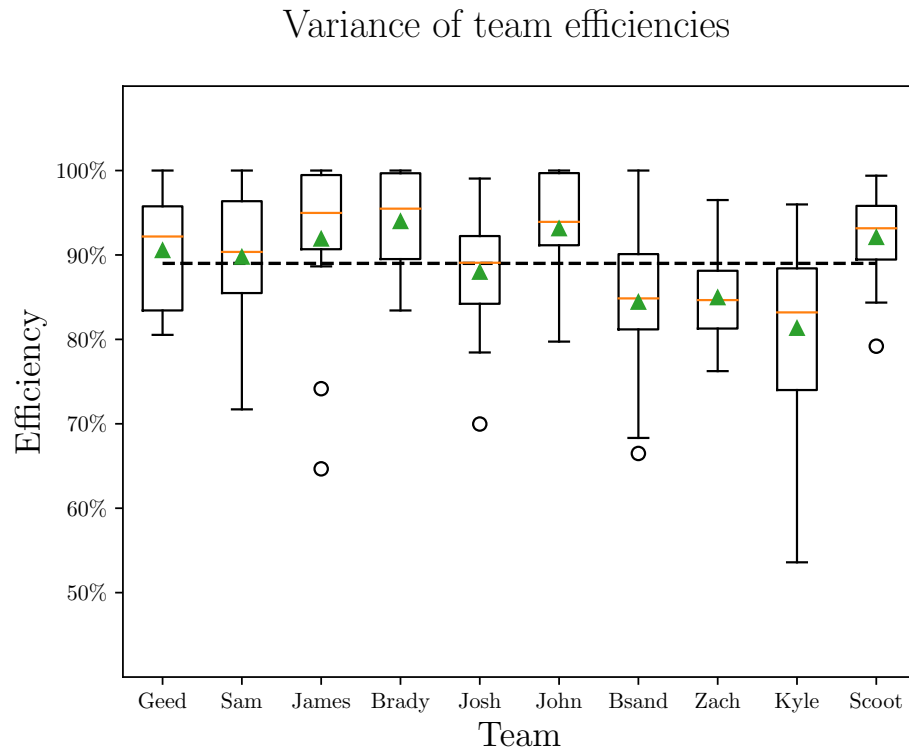
Figure 4.3: Variance of team owner efficiency over the duration of the season, where efficiency $= \frac{\text{Actual score}}{\text{Possible score}}$. Dashed black line indicates the league average efficiency.

5 Matchup point differential
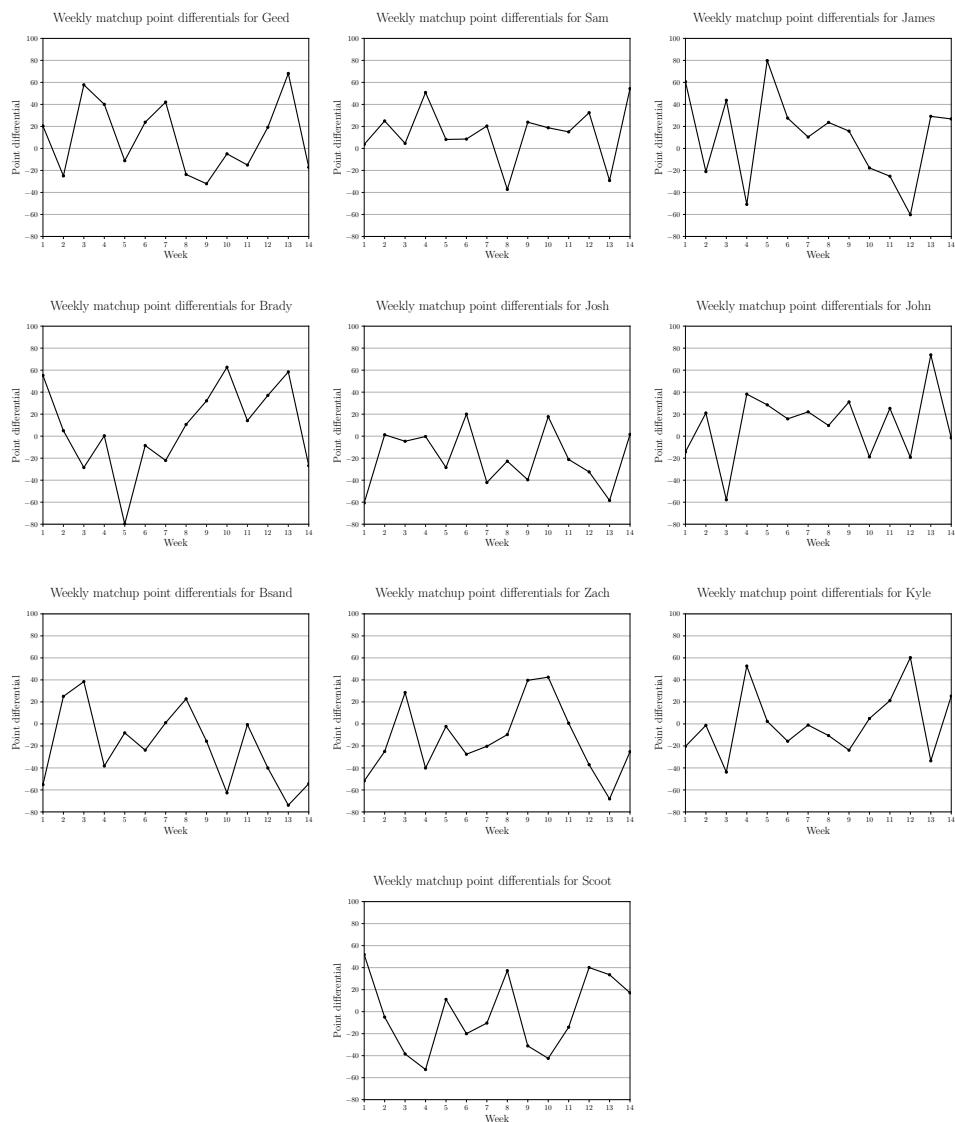
Figure 5.1: Point differentials in weekly matchups for each team.
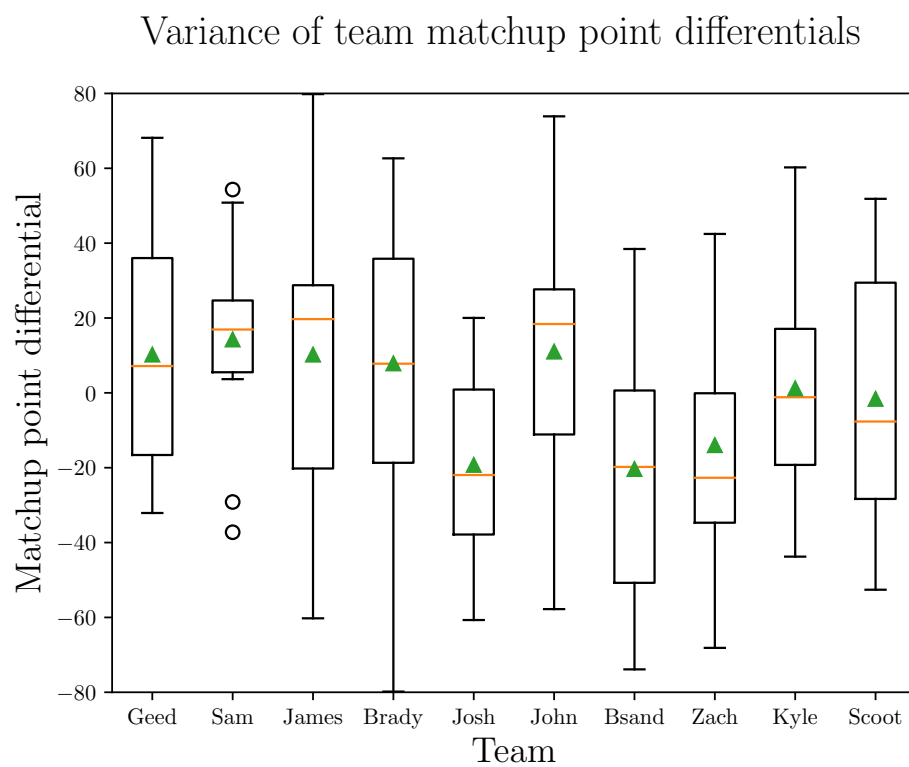
Figure 5.2: Variance of point differentials in weekly matchups over the duration of the season.

6   Regression analysis

In this section, the entire league's data is analyzed to determine whether or not a team's record is dependent upon one of the following "independent" variables:

- total points for (PF)

- coefficient of variation of points for (CV), defined as $\mathrm{CV} = \frac{\sigma_{\mathrm{dev}}}{\mu}$ where σ_{dev} is the standard deviation in PF and μ is the average PF

- a combination (interaction) of PF and CV (this is the PF:CV row in the tables)

as well as whether or not PF is correlated with CV.

To determine the aforementioned correlations, the `statsmodels` ordinary least-squares linear regression model was run with the dependent and independent variables as inputs. Recall that the fit of a model is given by a R^2 value, with higher R^2 generally indicating a better fit. Correlation between two variables is given by the "t-statistic", with larger values of t indicating positive correlations, and smaller values of t indicating negative correlations. For example, a large t-value between PF and record indicates that the dependent variable (record) is strongly correlated with the independent variable (PF). However, t-values mean little without a corresponding p-value (or probability of the "F-statistic"). If the p-value is less than, e.g., 0.05, then the probability that the independent variable is correlated with the dependent variable is greater than 0.95 or 95%. In some of the tables, there may be large t-values indicating strong correlation, but the corresponding p-values are also large, which tells us that the model cannot accurately predict strong correlations between dependent and independent variables. Evaluation of overall model fit when multiple independent variables are evaluated for correlation with a dependent variable is given by the R^2 value and the probability of the F-statistic.

It is also worth mentioning that the sample size of the data is small in that it only comprises the number of teams. In larger leagues with, e.g., 100 teams, the statistics models may produce better results. But for the smaller sample sizes considered herein, the kurtosis measurement (width of the "bell curve") provided by `statsmodels` should be evaluated with caution.

	coef	std err	t	P> |t|	[0.025	0.975]
Dep. Variable:	Record		**R-squared:**			0.495
Model:	OLS		**Adj. R-squared:**			0.432
Method:	Least Squares		**F-statistic:**			7.831
Date:	Sat, 30 Dec 2023		**Prob (F-statistic):**			0.0233
Time:	08:07:16		**Log-Likelihood:**			-20.212
No. Observations:	10		**AIC:**			44.42
Df Residuals:	8		**BIC:**			45.03
Df Model:	1					
Covariance Type:	nonrobust					

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-15.0610	7.910	-1.904	0.093	-33.301	3.179
PF	0.0145	0.005	2.798	0.023	0.003	0.026

Omnibus:	5.281	**Durbin-Watson:**		1.876
Prob(Omnibus):	0.071	**Jarque-Bera (JB):**		1.822
Skew:	0.976	**Prob(JB):**		0.402
Kurtosis:	3.748	**Cond. No.**		1.87e+04

Table 6.1: Ordinary least-squares regression analysis of correlation between team record and team total points (PF).

	coef	std err	t	P> |t|	[0.025	0.975]
Dep. Variable:	Record			**R-squared:**		0.042
Model:	OLS			**Adj. R-squared:**		-0.078
Method:	Least Squares			**F-statistic:**		0.3471
Date:	Sat, 30 Dec 2023			**Prob (F-statistic):**		0.572
Time:	08:07:16			**Log-Likelihood:**		-23.412
No. Observations:	10			**AIC:**		50.82
Df Residuals:	8			**BIC:**		51.43
Df Model:	1					
Covariance Type:	nonrobust					

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	10.1237	5.376	1.883	0.096	-2.273	22.520
CV	-16.0594	27.257	-0.589	0.572	-78.914	46.795

Omnibus:	0.684	**Durbin-Watson:**	1.665
Prob(Omnibus):	0.710	**Jarque-Bera (JB):**	0.574
Skew:	0.183	**Prob(JB):**	0.750
Kurtosis:	1.885	**Cond. No.**	31.8

Table 6.2: Ordinary least-squares regression analysis of correlation between team record and team points correlation of variation (CV).

	coef	std err	t	P> |t|	[0.025	0.975]
Dep. Variable:	Record			**R-squared:**		0.502
Model:	OLS			**Adj. R-squared:**		0.253
Method:	Least Squares			**F-statistic:**		2.014
Date:	Sat, 30 Dec 2023			**Prob (F-statistic):**		0.214
Time:	08:07:16			**Log-Likelihood:**		-20.141
No. Observations:	10			**AIC:**		48.28
Df Residuals:	6			**BIC:**		49.49
Df Model:	3					
Covariance Type:	nonrobust					

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-2.9118	60.175	-0.048	0.963	-150.154	144.330
PF	0.0073	0.038	0.192	0.854	-0.086	0.101
CV	-59.5169	300.600	-0.198	0.850	-795.058	676.025
PF:CV	0.0348	0.192	0.181	0.862	-0.436	0.505

Omnibus:	3.535	**Durbin-Watson:**	1.821
Prob(Omnibus):	0.171	**Jarque-Bera (JB):**	1.120
Skew:	0.795	**Prob(JB):**	0.571
Kurtosis:	3.403	**Cond. No.**	6.45e+05

Table 6.3: Ordinary least-squares regression analysis of correlation between team record and interaction between team total points (PF) and team correlation of variation of points (CV).

	coef	std err	t	P> |t|	[0.025	0.975]
Dep. Variable:	PF		R-squared:			0.039
Model:	OLS		Adj. R-squared:			-0.081
Method:	Least Squares		F-statistic:			0.3262
Date:	Sat, 30 Dec 2023		Prob (F-statistic):			0.584
Time:	08:07:16		Log-Likelihood:			-62.263
No. Observations:	10		AIC:			128.5
Df Residuals:	8		BIC:			129.1
Df Model:	1					
Covariance Type:	nonrobust					

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	1671.9976	261.627	6.391	0.000	1068.685	2275.311
CV	-757.6078	1326.552	-0.571	0.584	-3816.642	2301.427

Omnibus:	1.062	Durbin-Watson:	2.376
Prob(Omnibus):	0.588	Jarque-Bera (JB):	0.756
Skew:	-0.360	Prob(JB):	0.685
Kurtosis:	1.861	Cond. No.	31.8

Table 6.4: Ordinary least-squares regression analysis of correlation between team total points (CF) and team points correlation of variation (CV).

	coef	std err	t	P> |t|	[0.025	0.975]
Dep. Variable:	Record			**R-squared:**		0.673
Model:	OLS			**Adj. R-squared:**		0.509
Method:	Least Squares			**F-statistic:**		4.107
Date:	Sat, 30 Dec 2023			**Prob (F-statistic):**		0.0667
Time:	08:07:16			**Log-Likelihood:**		-24.008
No. Observations:	10			**AIC:**		56.02
Df Residuals:	6			**BIC:**		57.23
Df Model:	3					
Covariance Type:	nonrobust					

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-8.8364	88.581	-0.100	0.924	-225.587	207.914
PF	0.0151	0.056	0.267	0.798	-0.123	0.153
CV	-121.6379	442.504	-0.275	0.793	-1204.405	961.130
PF:CV	0.0788	0.283	0.278	0.790	-0.614	0.772

Omnibus:	1.845	**Durbin-Watson:**		1.636
Prob(Omnibus):	0.398	**Jarque-Bera (JB):**		0.729
Skew:	0.658	**Prob(JB):**		0.694
Kurtosis:	2.855	**Cond. No.**		6.45e+05

Table 6.5: Ordinary least-squares regression analysis of correlation between team record **using *median* wins** and interaction between team points and coefficient of variation of team points.