WHO WILL MAKE TEAM USA?

DAT10

OCTAVIA PAYNE

PROBLEM

105
TRYOUTS

35
SPOTS

DATA

DATA:

- 2020 observations (595 complete)
- Excluded mixed players (28)

TRAINING DATA:

- Statistical profile of 2013 world games players
    - 9 players
    - 68 observations (21 complete)

- Statistical profile of players on fury and riot
    - 97 players
    - 455 observations (126 complete)

HYPOTHESIS

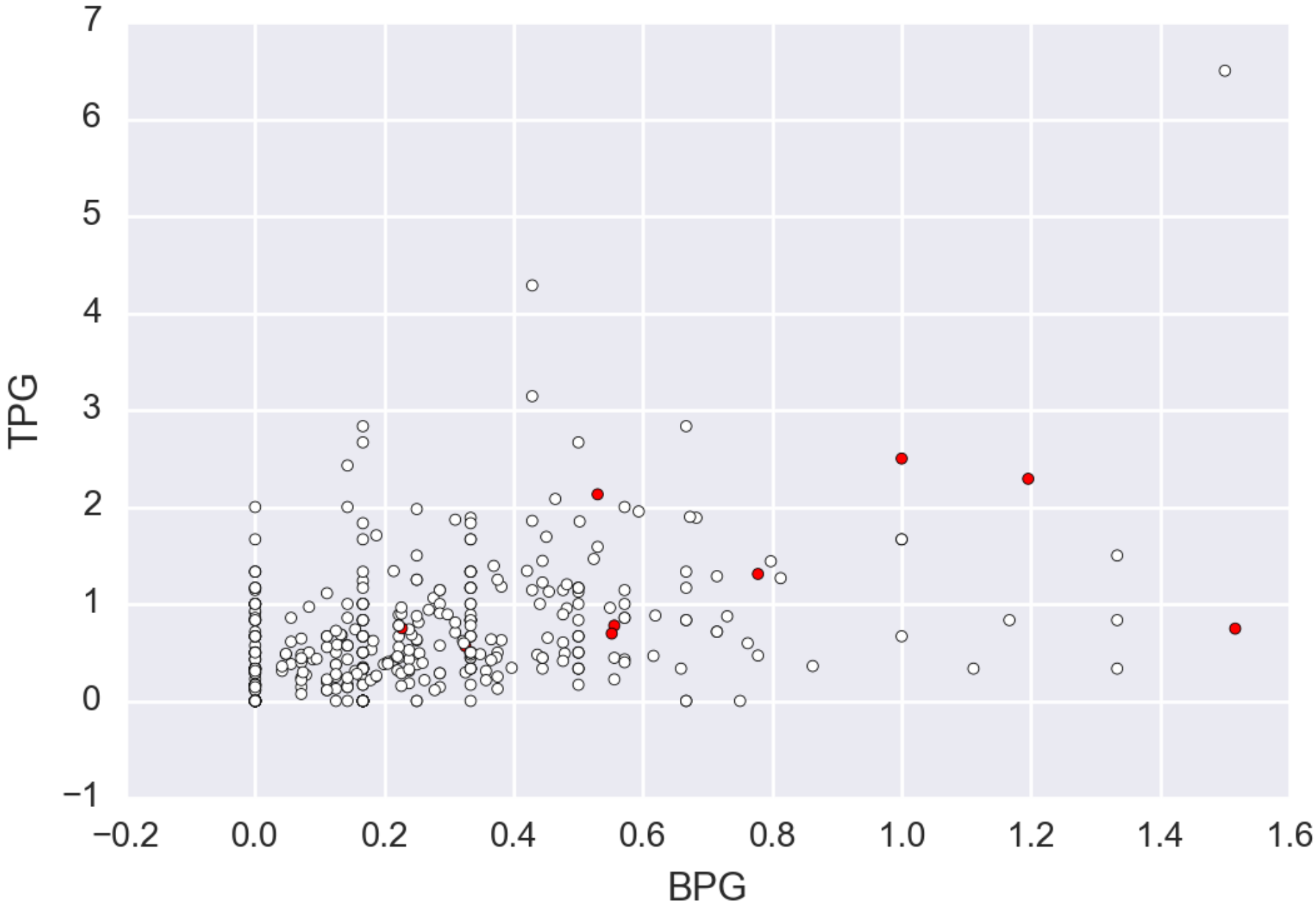
Players who are most consistent and demonstrate greatest sportsmanship have the best chances of making the roster

- Both coaches sub evenly (depth vs star-power)
- Their teams place heavy emphasis on team conduct

MODELING

METHODS:

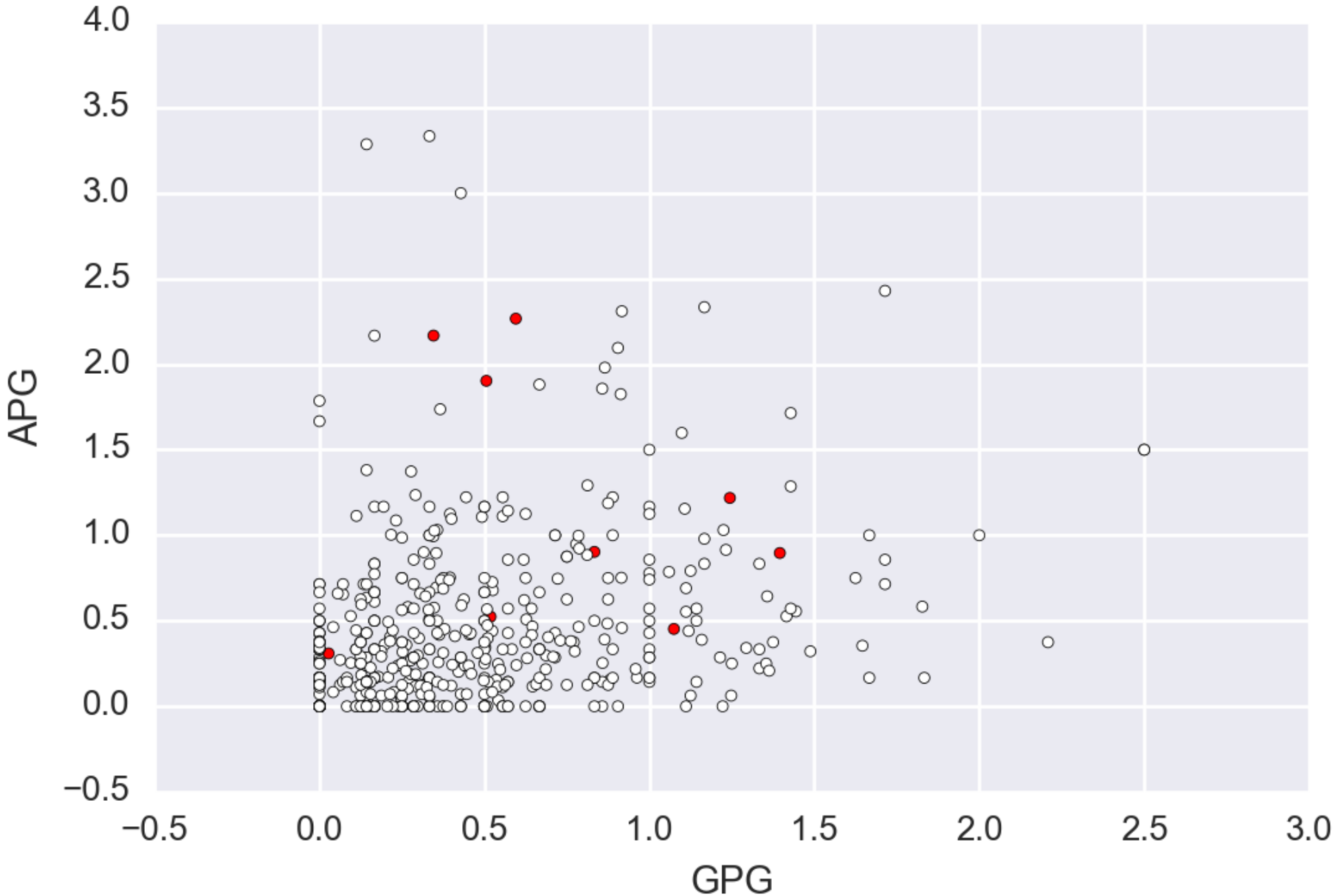
- KNN
- Clustering
- Decision Tree



Decision Tree: BPG vs TPG

1. Groupby('Name').mean()
2. Year <= 2013
3. Team != 'TEAM USA'

2.5



Decision Tree: GPG vs APG

1. Groupby('Name').mean()
2. Year <= 2013
3. Team != 'TEAM USA'

STANDARD DEVIATION OF FEATURES BY WG PLAYERS

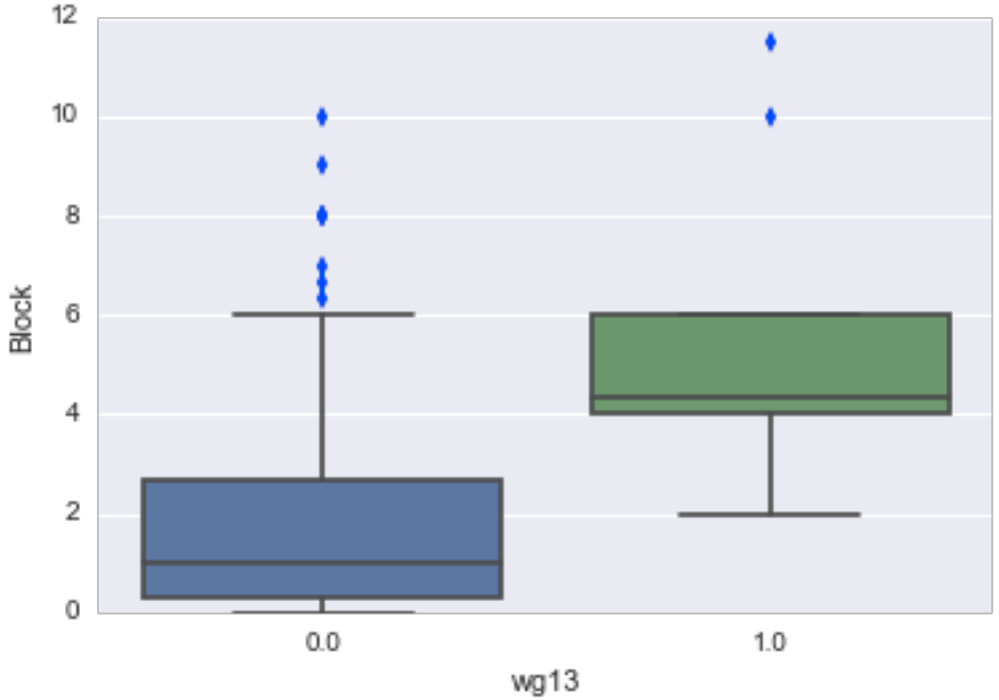
	STD %	PLAYER 1	PLAYER 2	PLAYER 3	PLAYER 4	PLAYER 5	PLAYER 6	PLAYER 7	PLAYER 8	PLAYER 9	AVG
GOAL	16%	-3%	41%	225%	-89%	225%	30%	257%	54%	149%	99%
GPG	10%	-20%	20%	149%	-91%	188%	17%	223%	38%	93%	68%
ASSIST	45%	584%	69%	44%	24%	213%	424%	137%	517%	182%	244%
APG	34%	418%	25%	8%	-1%	191%	355%	114%	442%	116%	186%
BLOCK	22%	245%	130%	44%	73%	149%	475%	561%	130%	245%	228%
BPG	16%	209%	121%	29%	35%	119%	375%	502%	110%	297%	200%
TURN	2%	116%	23%	-7%	106%	17%	291%	23%	243%	208%	113%
TPG	5%	84%	9%	-21%	58%	-2%	222%	5%	200%	252%	90%

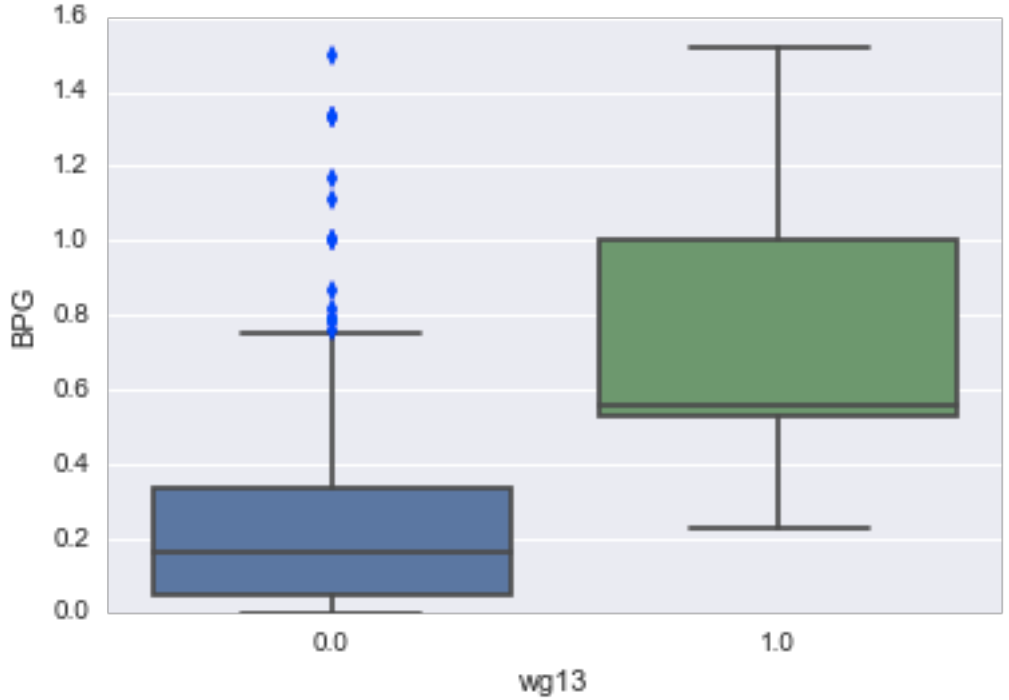
1. Every player is significantly more productive on offense and defense than rest of field
    - Green = significantly better than division avg.
    - Red = significantly worse than division avg.

2. Assists and Blocks seem to be where WG13 players stand out the most

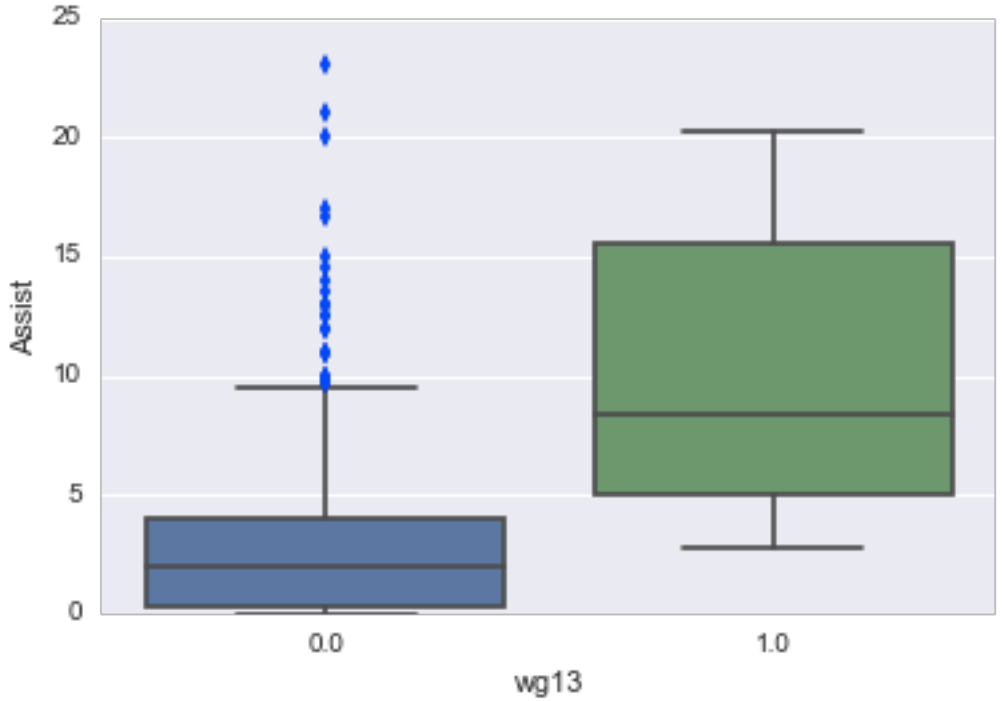
3. All but one player had significantly more turnovers (and in most cases turnovers per game) than the rest of the field
    - **Likely explanation:** these players touch the disc more often on their respective teams (no direct statistic to show)

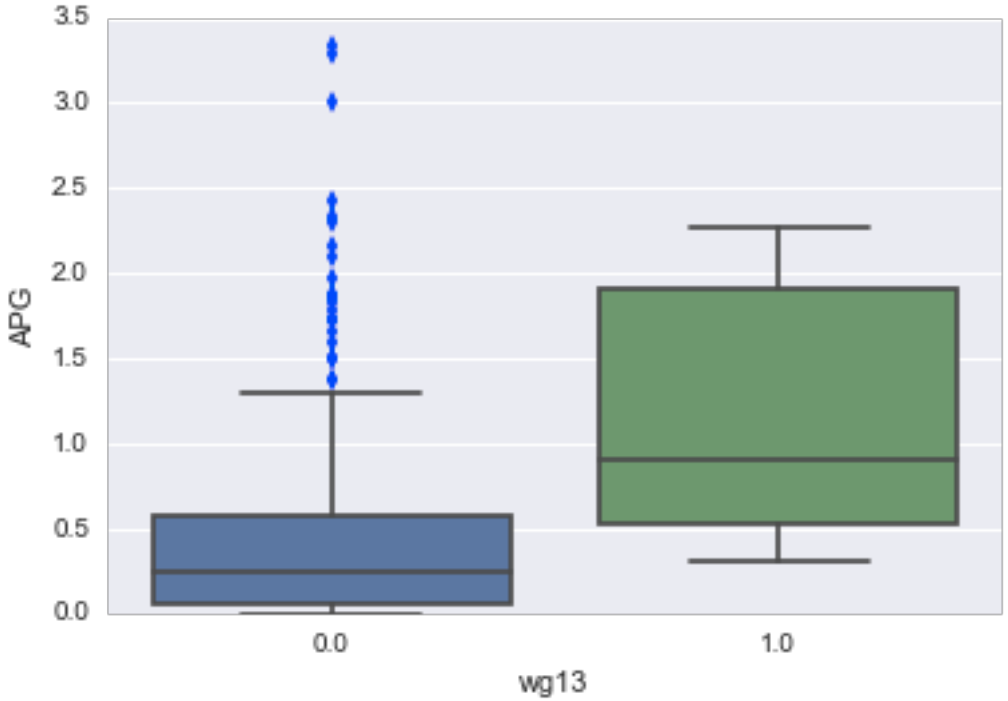
STATISTICAL SPREAD OF BLOCKS AND BLOCKS PER GAME (BPG)



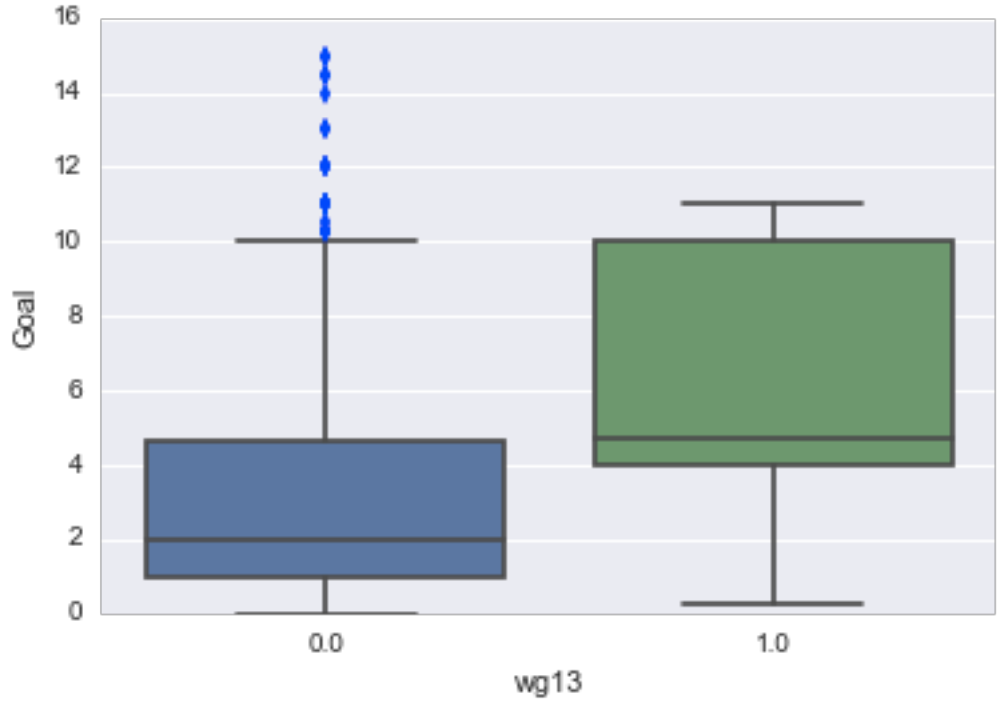


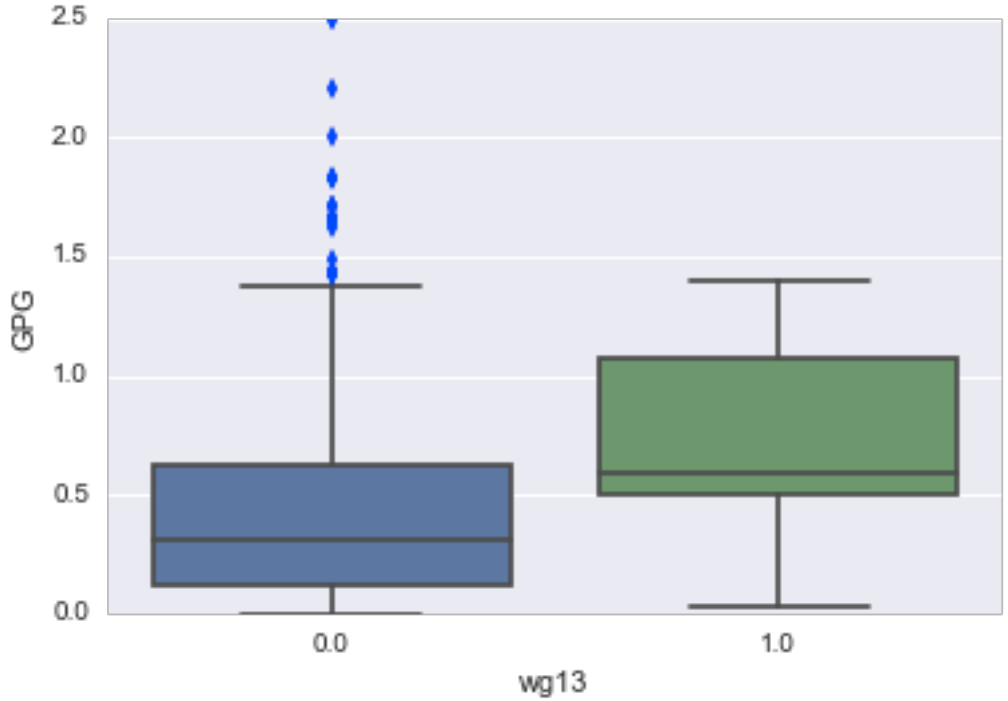
STATISTICAL SPREAD OF ASSISTS AND ASSISTS PER GAME (APG)





STATISTICAL SPREAD OF GOALS AND GOALS PER GAME (BPG)

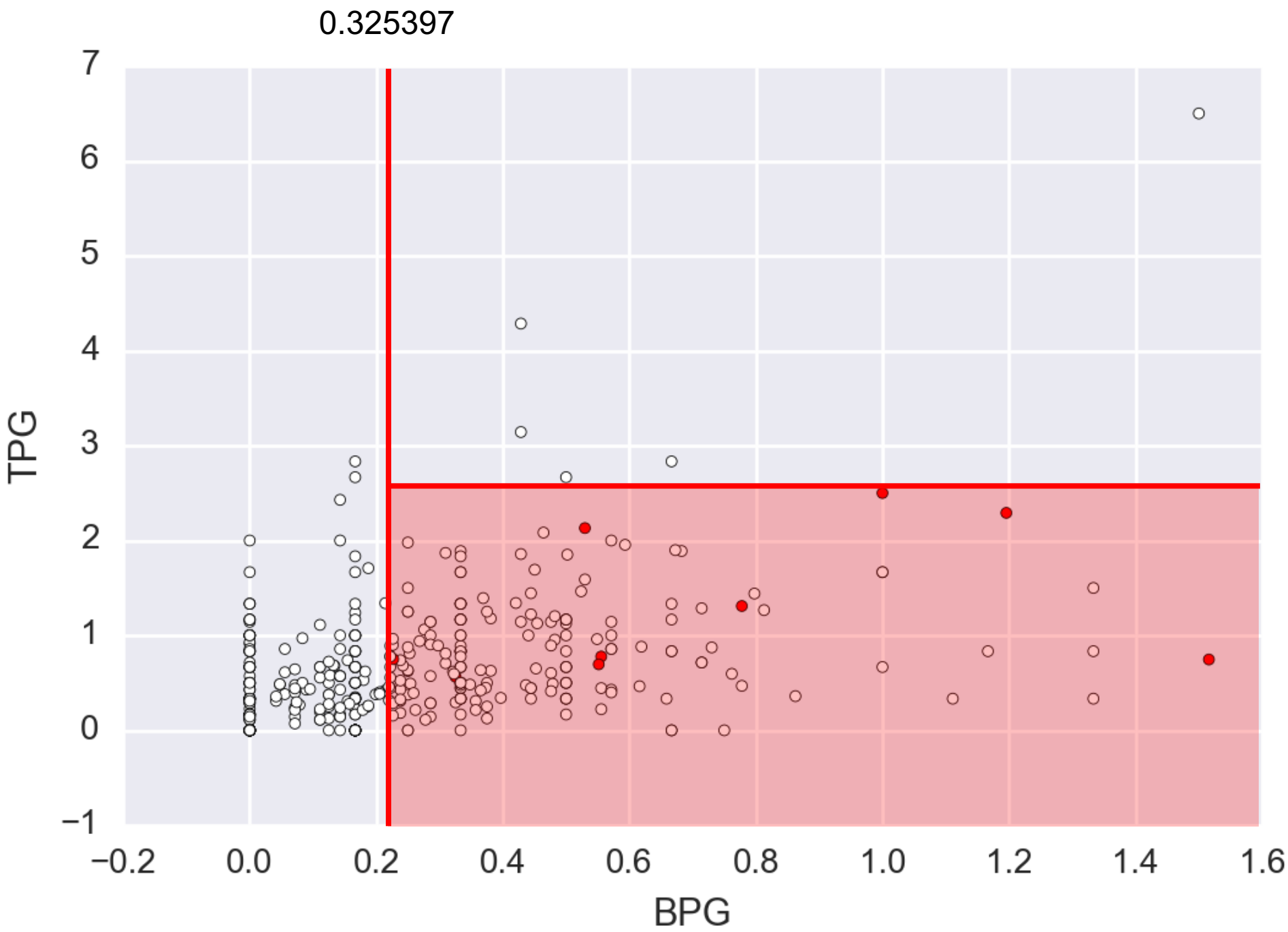




METHOD: DECISION TREE TUNED BY HAND

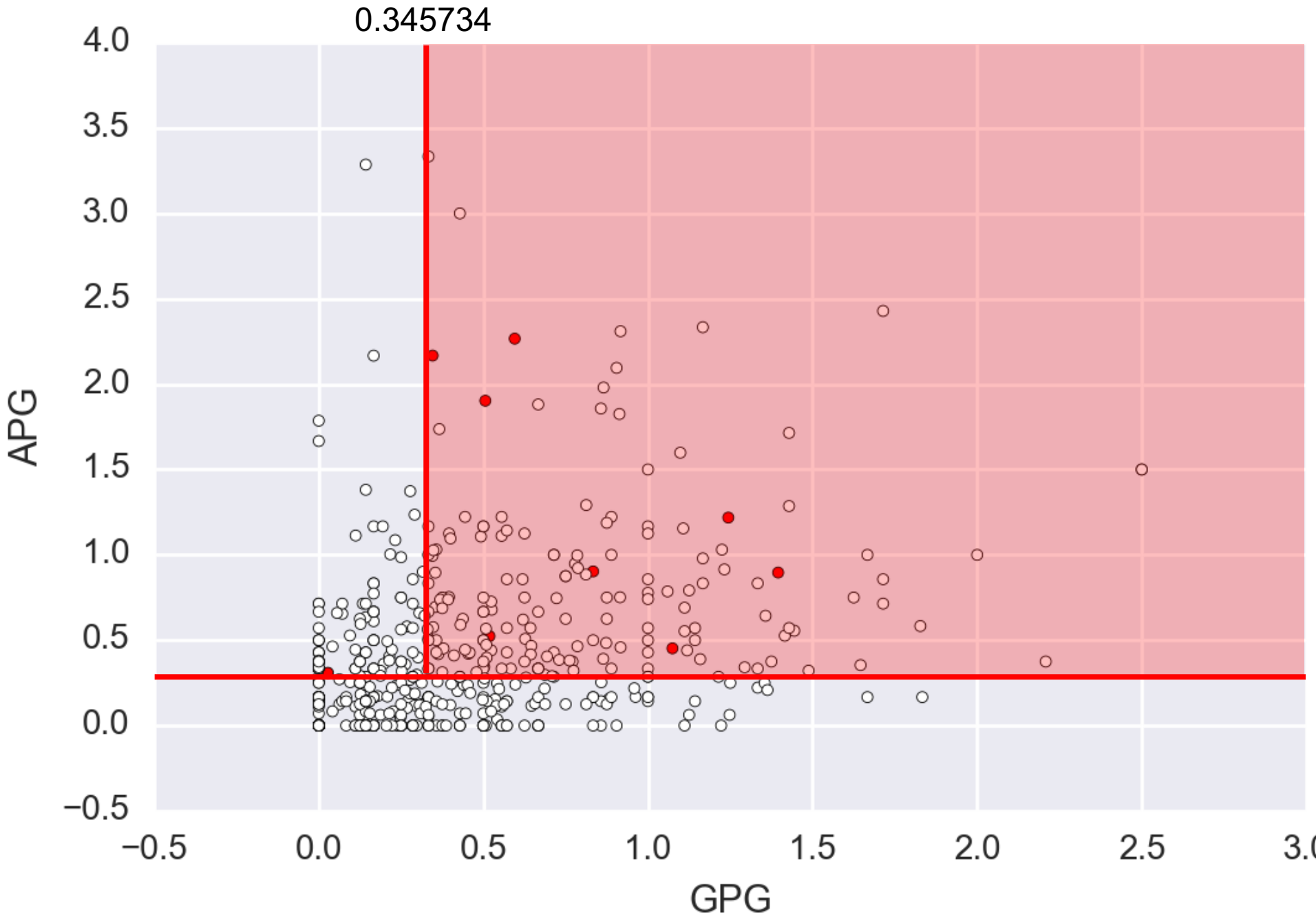
Feature	**Cutoff***
BPG	0.325397
APG	0.412037
GPG	0.345734

- Cutoffs determined by player with lowest mean observation, except for GPG, which excluded an outlier.
- Did not place an upper bound on BPG and APG nodes as higher BPG and APG are better.



Decision Tree: BPG vs TPG

1. Groupby('Name').mean()
2. Year <= 2013
3. Team != 'TEAM USA'



Decision Tree: GPG vs APG

1. Groupby('Name').mean()
2. Year <= 2013
3. Team != 'TEAM USA'

RESULTS

18 of 24 of women’s roster accurately predicted (75%)

21 of 35 of total accurately predicted (60%)

Sample size: 28 players from mixed division weren’t included in tryouts data

Prediction: 5 players from mixed division weren’t included in prediction

ACCURACY

n=77	**Predicted: Yes**	**Predicted: No**	
Actual: Yes	TP: 21	FN: 10	31
Actual: No	FP: 6	TN: 40	46
	27	50	n=77

Accuracy:
(TP +TN) / Total = (21+40) / 77 = 0.792

Error Rate:
(FP + FN) / Total = (6 + 10) / 77 = 0.21

Sensitivity (TPR):
TP / Actual Yes = 21 / 31 = 0.677

Specificity:
TN / Actual No = 40 /46 = 0.869

- **Sample size:** 28 players from mixed division weren't included in data scrape
- **Prediction:** 5 players from mixed division weren't included in data scrape

HURDLES/TAKEAWAYS

- Poor stats (quantity, type, and consistency)
  - Need stats at every tournament
  - Add yards gained, conceded, touches, hockey assist, points played
- Difficult to include mixed players in automated fashion (4 WUGC players from mixed)
- Name changes
- FURY and RIOT not significantly different from other teams
- Spirit scores did not impact

THANK YOU!

DAT10

OCTAVIA PAYNE