Olympic Body Type Analysis

IST 687 – Kyle Wojtaszek

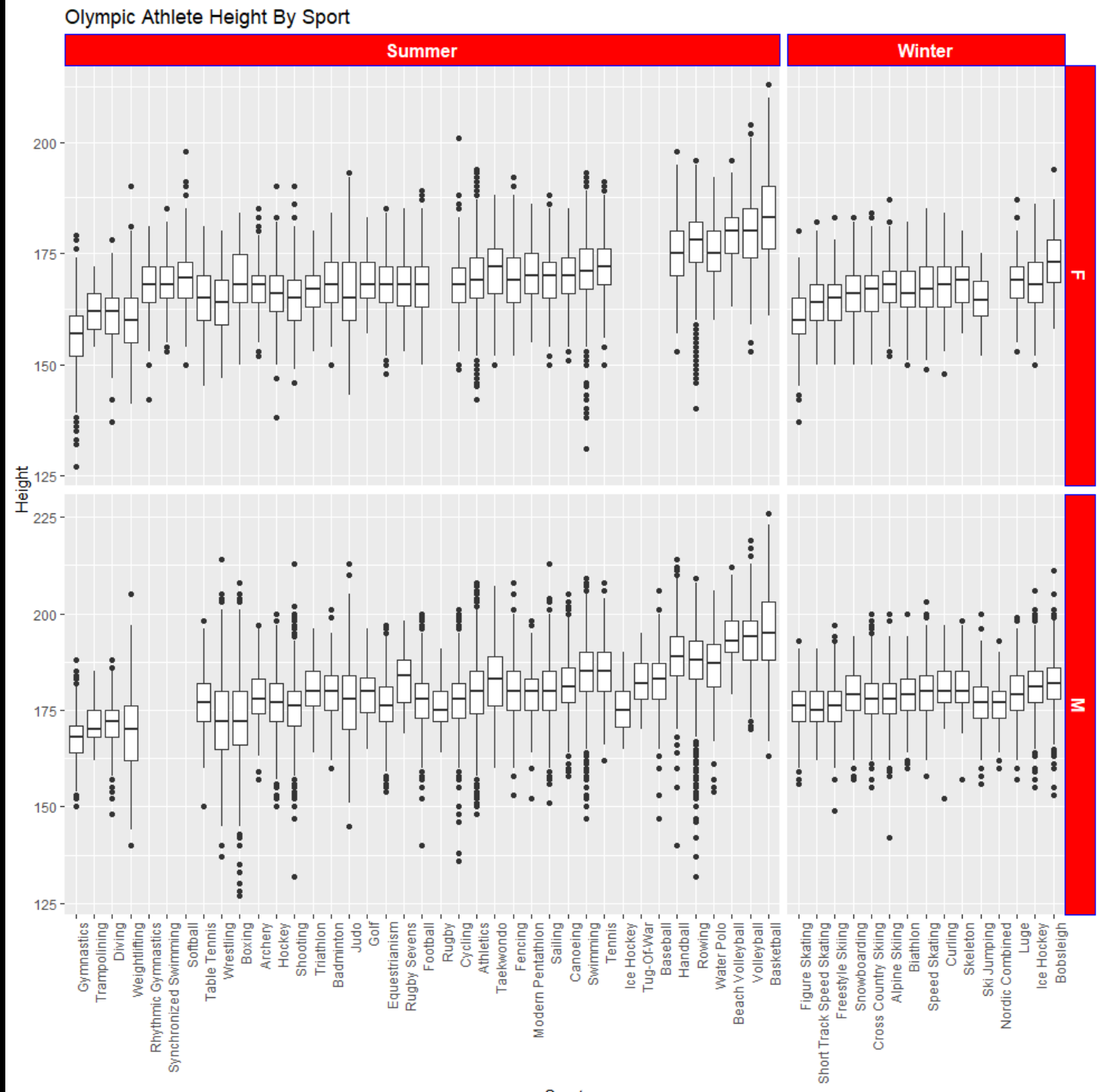
QUESTIONS

- Variation of body type per sport
  - Which sports have the highest variation?
  - Lowest?
  - Do the sports with highest/lowest variation share characteristics?

- Prediction of most likely sport by body type
  - Can we predict with reasonable accuracy the best sports for an individual given their physical characteristics?

HEIGHT

- Overview
  - Lots of similar distributions in the middle of the graph
  - Summer games have more sports with higher weights
  - Winter games have fewer outliers
  - Females generally about 10cm shorter in sports which are contested with both genders



HEIGHT – HIGHEST VARIATION

Sport	Sex	25thPerc	Median	75thPerc	Count	Quartile Range
Basketball	M	188.0	195.0	203.0	2630	15.0
Wrestling	M	165.0	172.0	180.0	5042	15.0
Basketball	F	176.0	183.0	190.0	1251	14.0
Boxing	M	166.0	172.0	180.0	4508	14.0
Judo	M	170.0	178.0	184.0	2400	14.0
Weightlifting	M	162.0	170.0	176.0	2548	14.0
Judo	F	160.0	165.0	173.0	1000	13.0
Taekwondo	M	176.0	183.0	189.0	306	13.0
Boxing	F	164.0	168.0	175.0	70	11.0
Volleyball	F	174.0	180.0	185.0	1495	11.0
Athletics	M	174.0	180.0	185.0	21882	11.0
Rugby Sevens	M	177.0	184.0	188.0	151	11.0
Water Polo	M	181.0	187.0	192.0	2300	11.0

- Basketball
  - Small guards vs. large centers
- Combat Sports
  - Wrestling, Boxing, Judo, Taekwondo
  - Weight classes encourage large variation of competitors (in both height and weight)
- Weightlifting also has weight classes

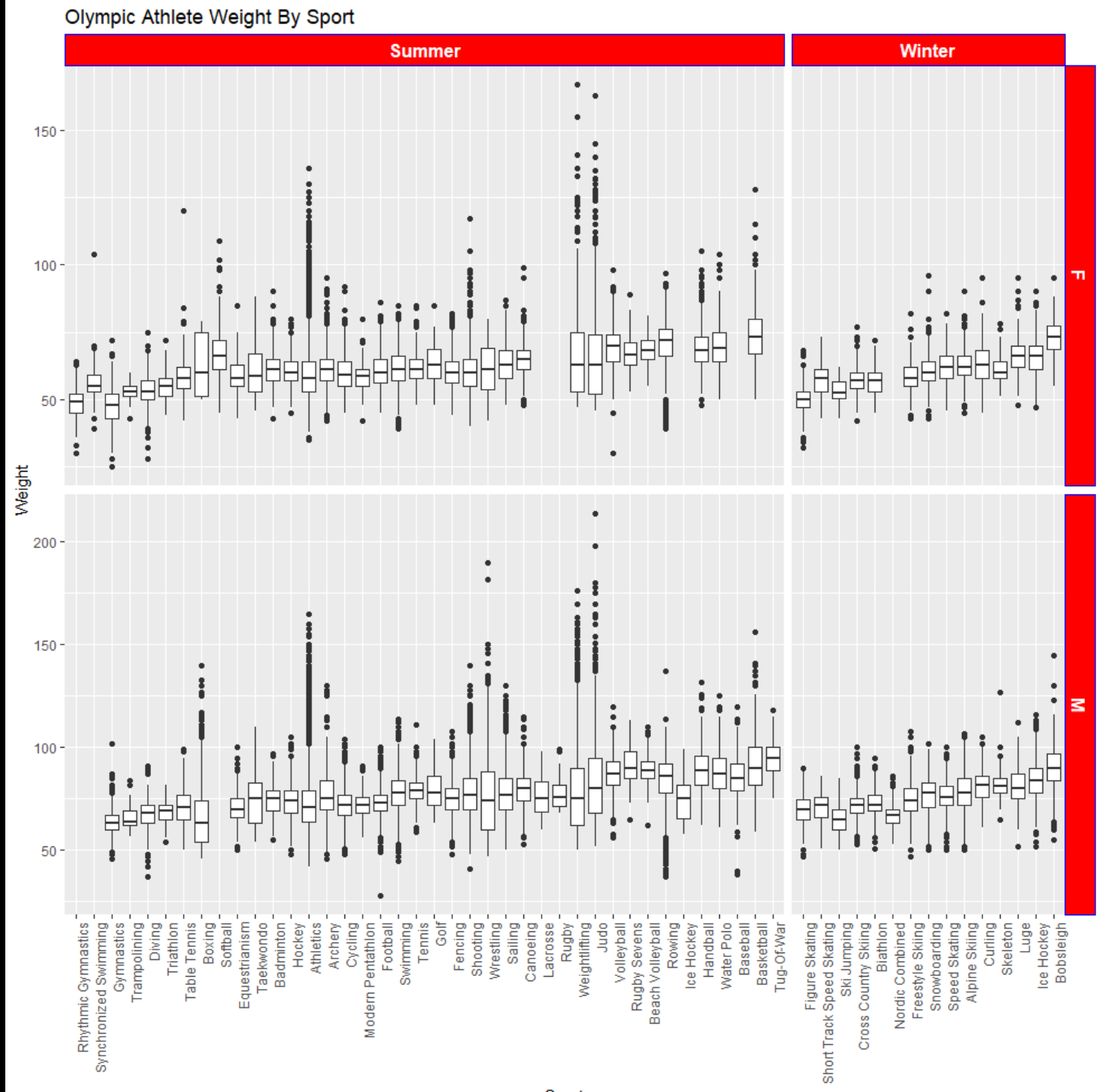
HEIGHT – LOWEST VARIATION

Sport	Sex	25thPerc	Median	75thPerc	Count	Quartile Range
Archery	F	164.0	168.0	170.0	907	6.0
Alpine Skiing	F	164.0	168.0	171.0	2638	7.0
Luge	F	165.0	169.0	172.0	351	7.0
Synchronized Swimming	F	165.0	168.0	172.0	864	7.0
Triathlon	F	163.0	167.0	170.0	262	7.0
Diving	M	168.0	172.0	175.0	1093	7.0
Gymnastics	M	164.0	168.0	171.0	11022	7.0
Nordic Combined	M	173.0	177.0	180.0	1120	7.0
Trampolining	M	168.0	170.0	175.0	76	7.0

- Top 5 all female events
- Sports where being short is an advantage
  - Alpine Skiing
    - Lower center of gravity
  - Luge
    - Less weight
  - Synchronized Swimming, Diving, Gymnastics, Trampolining
    - Easier to twist/flip
- Archery is a bit of a mystery

WEIGHT

- Overview
  - Similar to height, lots of similar sports in the middle of the graph
  - Much larger amount of outliers, especially in Summer games
  - Females generally 10-15kg lighter across sports



WEIGHT – HIGHEST VARIATION

Sport	Sex	25thPerc	Median	75thPerc	Count	Quartile Range
Weightlifting	M	62.0	75.0	90.00	3340	28.00
Wrestling	M	60.0	74.0	88.00	5002	28.00
Judo	M	68.0	80.0	95.00	2441	27.00
Boxing	F	51.0	60.0	75.00	61	24.00
Judo	F	52.0	63.0	74.00	1005	22.00
Weightlifting	F	53.0	63.0	75.00	463	22.00
Boxing	M	54.0	63.0	74.00	4489	20.00
Taekwondo	M	63.0	75.0	83.00	306	20.00
Basketball	M	82.0	90.0	100.00	2471	18.00
Wrestling	F	53.5	61.0	69.00	303	15.50

- Weight classes show up again
  - Weightlifting, Wrestling, Judo, Boxing, Taekwondo
  - Encourage large variation in weights
- Basketball
  - Large variation in height leads to variation in weight as well

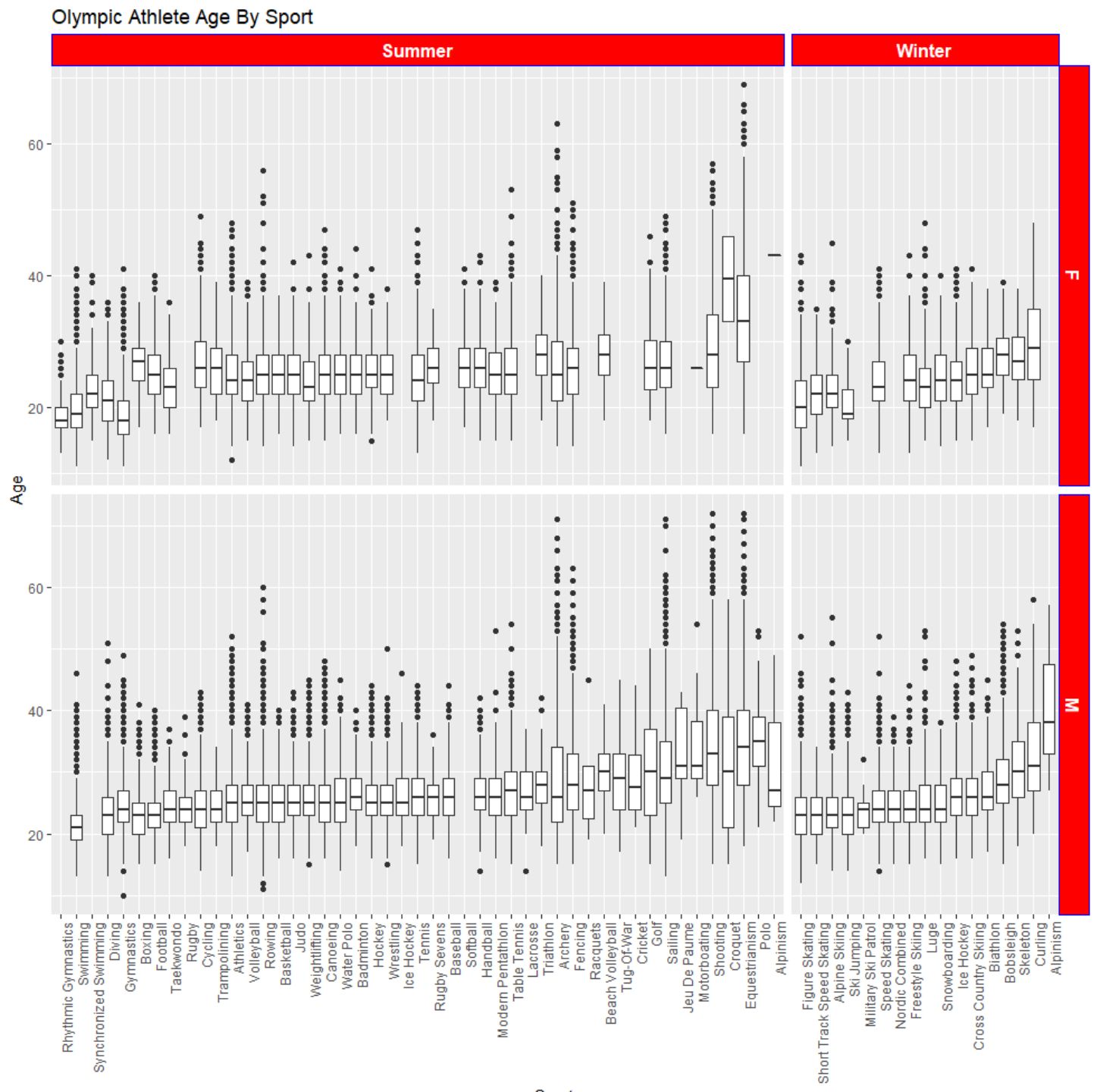
WEIGHT – LOWEST VARIATION

Sport	Sex	25thPerc	Median	75thPerc	Count	Quartile Range
Trampolining	F	51.0	53.0	55.00	75	4.00
Cross Country Skiing	F	54.0	57.0	60.00	3088	6.00
Figure Skating	F	47.0	50.0	53.00	770	6.00
Modern Pentathlon	F	55.0	58.5	61.00	164	6.00
Skeleton	F	58.0	60.0	64.00	65	6.00
Synchronized Swimming	F	53.0	55.0	59.00	850	6.00
Alpine Skiing	F	59.0	62.0	66.00	2605	7.00
Beach Volleyball	F	65.0	68.0	72.00	265	7.00

- Similar to height list
- Sports in which being lighter is an advantage
  - Trampolining, Figure Skating, Synchronize Swimming
    - Easier to flip/spin
  - Cross Country Skiing, Modern Pentathlon, Alpine Skiing, Skeleton
    - Carrying less weight is faster
- Lack of variation is most likely due to limits on the human body.
  - Athletes can only be so light before it becomes unhealthy

AGE

- Overview
  - Once again, middle of graph shows relatively low variation
    - Mid 20's is generally human's peak athletic years
  - A lot more higher outliers than lower outliers
    - Early Olympics did not have any many specialized athletic programs as today. Allowed older athletes to stay elite longer.



AGE – HIGHEST VARIATION

Sport	Sex	25thPerc	Median	75thPerc	Count	Quartile Range
Croquet	M	21.0	30.0	39.0	11	18.0
Alpinism	M	33.0	37.0	48.0	15	15.0
Golf	M	23.0	30.0	37.0	167	14.0
Croquet	F	33.0	39.5	46.0	4	13.0
Equestrianism	F	27.0	33.0	40.0	1244	13.0
Archery	M	22.0	26.0	34.0	1249	12.0
Equestrianism	M	28.0	34.0	40.0	4907	12.0
Shooting	M	28.0	33.0	40.0	8905	12.0

- Generally events which are not traditionally associated with athletics abilities such as cardiovascular endurance or raw strength
- Without requiring cardio or strength, easier for competitors to participate for a longer range of their lifetime

AGE – LOWEST VARIATION

Sport	Sex	25thPerc	Median	75thPerc	Count	Quartile Range
Rhythmic Gymnastics	F	17.0	18.0	20.0	658	3.0
Football	M	21.0	23.0	25.0	5358	4.0
Military Ski Patrol	M	21.0	24.0	25.0	19	4.0
Rugby	M	22.0	24.0	26.0	114	4.0
Swimming	M	19.0	21.0	23.0	12910	4.0

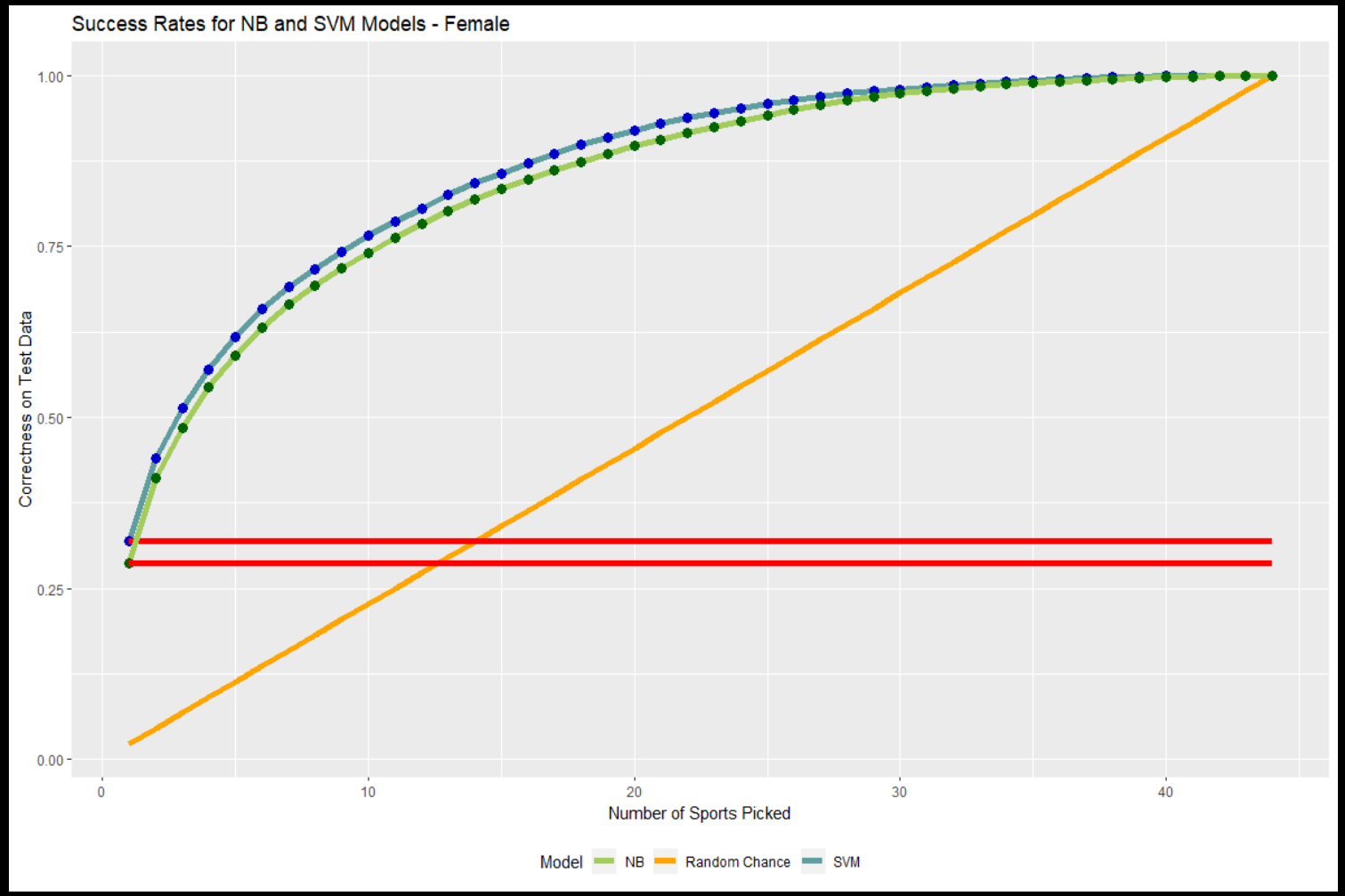
- Majority of sports have a 5 year quartile range, so very few sports with 'low' variation
- Football has an age cap of 23 years old in recent games, which artificially squeezes the age distribution

PREDICTION OF SPORTS - SETUP

- Predict best sports based on age, height, and weight
  - Only data from 1970 onward to capture modern sports and modern training techniques
  - Should we include Country as well?
- Separate models for males and females
  - Body types are different, no need to include extra variation in our model by combining both.
- Explore a couple different methods for prediction
- Our model will pick a Top N sports for a given body. Whoever is requesting the data can choose N.

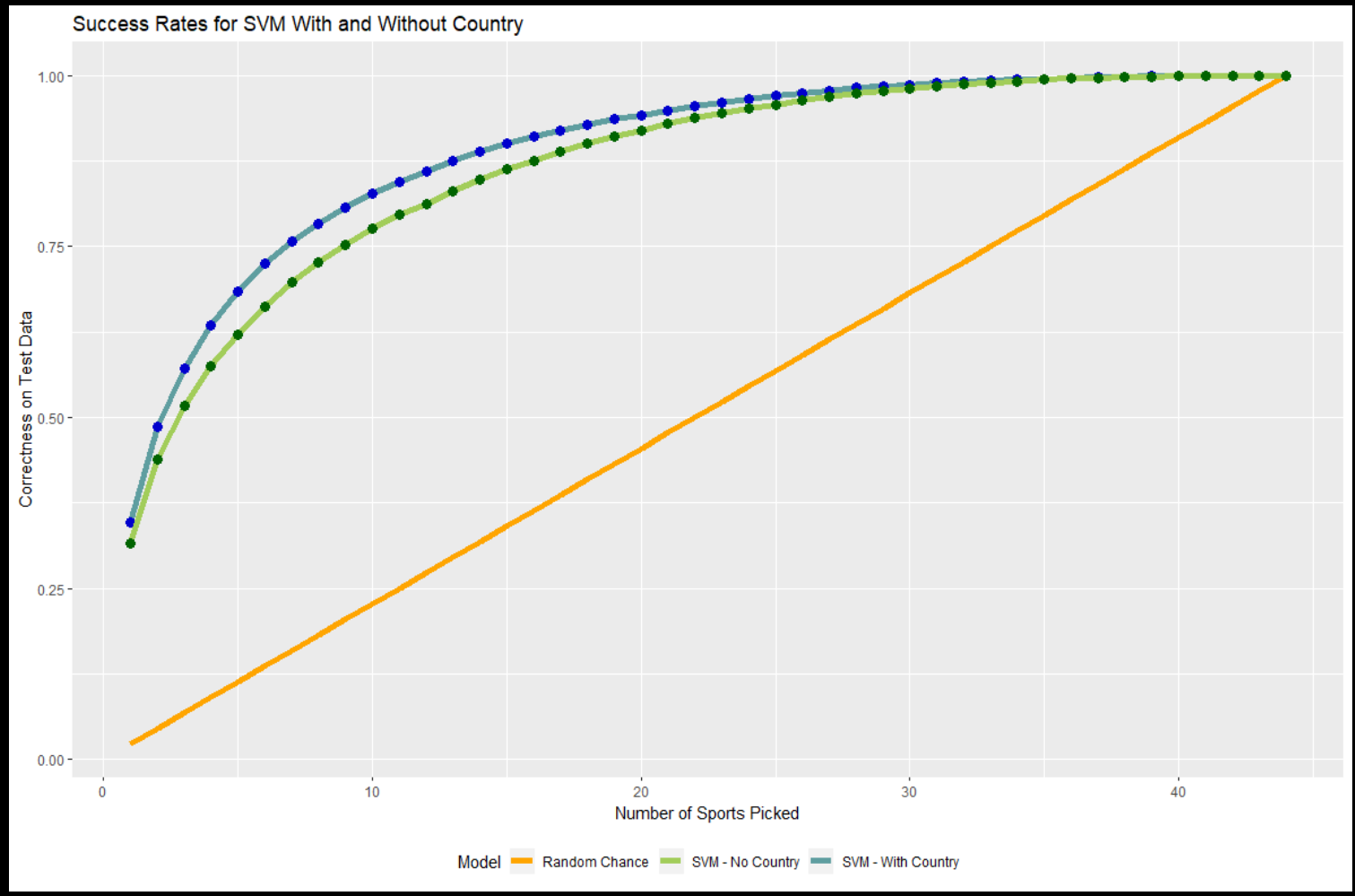
PREDICTION OF SPORT – COMPARISON OF MODELS

- Female model pictures, male model is similar
- Compare two models (SVM and NB)
- Both models significantly outperform random chance
- SVM model is consistently ~2% better performing than NB regardless of how many sports we pick



PREDICTION OF SPORT: SHOULD WE INCLUDE COUNTRY?

- Adding country to the model increases accuracy by ~2% across both genders
- Additionally, it increases accuracy ~5% in the first 10 sports picked, as shown in the graph
- Overall, a worthwhile addition to the model, small price for a big gain in accuracy



PREDICTION OF SPORT – CONCLUSIONS

- We can use our model to predict a sport for a given body type with reasonable accuracy
  - ~70% if we pick five sports
  - ~80% if we pick ten sports
- Using country was a big boon for our prediction
- Male and female prediction accuracy is basically the same