Title: Derby Point Standings 2020
Author: [redacted]
Date: 10/8/2021

Part 1

Introduction

The Thoroughbred racing season of 2020 was much different than any other. I've never really followed the races as much as I'd like being such a big horse fan. After events being put to a halt due to the recent pandemic. I decided to take this time to look into the past racing season and get caught up on the standings of 2020. My dataset can be found at www.bloodhorse.com. I also gathered knowledge of the file from Dr. Nicholas Jacob.

#	Horse	Parentage	Points	Non-Restricted Stakes Earnings	Career Earnings	Trainer	Breeder	Owner
1	Tiz the Law	Constitution—Tizfiz	372	$2,572,400	$2,615,300	Barclay Tagg	Twin Creeks Farm	Sackatoga Stable
2	Authentic	Into Mischief—Flawless	200	$2,840,000	$2,871,200	Bob Baffert	Peter E. Blum Thoroughbreds, LLC	SF Racing LLC, Starlight Racing, Madaket Stables LLC, Hertrich, III, Frederick, Fielding, John D. and Golconda Stables
3	Art Collector	Bernardini—Distorted Legacy	150	$476,461	$664,380	Joe Sharp	W. Bruce Lunsford	Bruce Lunsford

4	Honor A. P.	Honor Code—Hollywood Story	140	$490,000	$532,200	John Shirreffs	George Krikorian	C R K Stable LLC
5	Ny Traffic	Cross Traffic—Mamie Reilly	110	$472,820	$565,470	Saffie Joseph, Jr.	Brian Culnan	Fanelli, John, Cash is King LLC, LC Racing and Braverman, Paul
6	King Guillermo	Uncle Mo—Slow Sand	90	$317,050	$340,350	Juan Carlos Avila	Carhue Investments, Grouseridge Ltd. & Marengo Investments	Victoria's Ranch
7	Thousand Words	Pioneerof the Nile—Pomeroys Pistol	83	$297,000	$327,000	Bob Baffert	Hardacre Farm	Jerome S. Albaugh Family Stables LLC and Spendthrift Farm LLC
8	Dr Post	Quality Road—Mary Delaney	80	$340,035	$370,635	Todd Pletcher	Cloyce C. Clark	St. Elias Stable
9	Max Player	Honor Code—Fools in Love	60	$427,500	$463,500	Linda Rice	K & G Stables	George E. Hall
10	Caracaro	Uncle Mo—Peace Time	60	$205,000	$238,800	Gustavo Delgado	SF Bloodstock LLC	Global Thoroughbred and Top Racing, LLC
11	Country Grammer	Tonalist—Arabian Song	50	$106,400	$157,320	Chad Brown	Scott Pierce &	Paul P. Pompa, Jr.

							Debbie Pierce	
12	Pneumatic	Uncle Mo—Teardrop	45	$169,250	$256,850	Steve Asmussen	Winchell Thoroughbreds	Winchell Thoroughbreds
13	Enforceable	Tapit—Justwhistledixie	43	$314,550	$397,150	Mark Casse	Clearsky Farms	John C. Oxley
14	Swiss Skydiver	Daredevil—Expo Gold	40	$1,141,820	$1,192,980	Kenny McPeek	WinStar Farm	Peter J. Callahan
15	Shivaree	Awesome of Course—Garter Belt	40	$311,005	$370,505	Ralph Nicks	Jacks or Better Farm Inc.	Jacks or Better Farm, Inc.
16	Major Fed	Ghostzapper—Bobby's Babe	38	$179,100	$215,600	Gregory Foley	Lloyd Madison IV, LLC	Lloyd Madison Farms, IV LLC
17	Storm the Court	Court Vision—My Tejana Storm	36	$1,273,851	$1,310,451	Peter Eurton	Stepping Stone Farm	Exline-Border Racing LLC, Bernsen, David A., Wilson, Susanna and Hudock, Dan
18	Attachment Rate	Hard Spun—Aristra	35	$108,525	$143,732	Dale Romans	Mr. & Mrs. C. Oliver Iselin III	Bakke, Jim and Isbister, Gerald
19	Anneau d'Or	Medaglia d'Oro—Walk Close	32	$435,821	$453,821	Blaine Wright	Highland Yard LLC	Peter Redekop B. C., Ltd.
20	Sole Volante	Karakontie—Light Blow	30	$273,510	$323,310	Patrick Biancone	Flaxman Holdings Limited	Biancone, Andie and Limelight Stables Corp.

21	Finnick the Fierce	Dialed In—Southern Classic	25	$121,700	$191,290	Rey Hernandez	Paige Jillian Blu Sky Stables	Monge, Arnaldo and Hernandez, Rey
22	Uncle Chuck	Uncle Mo—Forest Music	20	$120,000	$150,000	Bob Baffert	Stonestreet Thoroughbred Holdings	Karl Watson, Michael E. Pegram and Paul Weitman
23	Candy Tycoon	Twirling Candy—Liberty's Lyric	20	$84,250	$169,850	Todd Pletcher	Jerry Romans Jr.	Mathis Stable LLC
24	Winning Impression	Paynter—Unbridled Sonya	20	$54,822	$98,552	Dallas Stewart	WinStar Farm	West Point Thoroughbreds and Pearl Racing
25	Shotski	Blame—She Cat	19	$212,466	$236,222	Jeremiah O'Dwyer	Springland Farm & Prime Bloodstock LLC	Wachtel Stable, Barber, Gary, Pantofel Stable and Karty, Mike
26	South Bend	Algorithms—Sandra's Rose	18	$291,902	$390,114	Stanley Hough	Highclere, Inc.	Sagamore Farm LLC
27	Necker Island	Hard Spun—Jenny's Rocket	14	$74,808	$199,730	Stanley Hough	Stonestreet Thoroughbred Holdings LLC	Sagamore Farm LLC and Hough, Stanley M.
28	Rowdy Yates	Morning Line—Spring Station	7	$148,908	$346,556	Steven Asmussen	Tracy Rene Strachan	L and N Racing LLC
29	Cezanne	Curlin—Achieving	5	$6,000	$63,000	Bob Baffert	Hill 'n' Dale	Mrs. John Magnier,

							Equine Holdings, Inc. & St. Elias Stables, LLC	Michael Tabor, Derrick Smith and St. Elias Stable
30	Super John	Super Saver—Atlantic Park	2	$9,500	$72,560	John Servis	Fred W. Hertrich III & John D. Fielding	Ryoji Green

Data and Variable Overview

This data is relatively well put together, but to avoid any confusion. I will mainly be focusing on seven out of nine characteristics of the data. This does not include the ‘Non-Restricted Stakes Earnings’ or the individual ‘Owners.’ I will analyze two categorical variables, which are Trainers and Breeders. They are both nominal variables. Next, I’ll be looking at quantitative variables. Both Ranking and Points are labeled as discrete, while Career Earnings is continuous.

#	Horse	Parentage	Points	Career Earnings	Trainer	Breeder
1	Tiz the Law	Constitution—Tizfiz	372	$2,615,300	Barclay Tagg	Twin Creeks Farm
2	Authentic	Into Mischief—Flawless	200	$2,871,200	Bob Baffert	Peter E. Blum Thoroughbreds, LLC
3	Art Collector	Bernardini—Distorted Legacy	150	$664,380	Joe Sharp	W. Bruce Lunsford
4	Honor A. P.	Honor Code—Hollywood Story	140	$532,200	John Shirreffs	George Krikorian
5	Ny Traffic	Cross Traffic—Mamie Reilly	110	$565,470	Saffie Joseph, Jr.	Brian Culnan
6	King Guillermo	Uncle Mo—Slow Sand	90	$340,350	Juan Carlos Avila	Carhue Investments, Grouseridge Ltd. & Marengo Investments
7	Thousand Words	Pioneerof the Nile—Pomeroys Pistol	83	$327,000	Bob Baffert	Hardacre Farm

8	Dr Post	Quality Road—Mary Delaney	80	$370,635	Todd Pletcher	Cloyce C. Clark
9	Max Player	Honor Code—Fools in Love	60	$463,500	Linda Rice	K & G Stables
10	Caracaro	Uncle Mo—Peace Time	60	$238,800	Gustavo Delgado	SF Bloodstock LLC
11	Country Grammer	Tonalist—Arabian Song	50	$157,320	Chad Brown	Scott Pierce & Debbie Pierce
12	Pneumatic	Uncle Mo—Teardrop	45	$256,850	Steve Asmussen	Winchell Thoroughbreds
13	Enforceable	Tapit—Justwhistledixie	43	$397,150	Mark Casse	Clearsky Farms
14	Swiss Skydiver	Daredevil—Expo Gold	40	$1,192,980	Kenny McPeek	WinStar Farm
15	Shivaree	Awesome of Course—Garter Belt	40	$370,505	Ralph Nicks	Jacks or Better Farm Inc.
16	Major Fed	Ghostzapper—Bobby's Babe	38	$215,600	Gregory Foley	Lloyd Madison IV, LLC
17	Storm the Court	Court Vision—My Tejana Storm	36	$1,310,451	Peter Eurton	Stepping Stone Farm
18	Attachment Rate	Hard Spun—Aristra	35	$143,732	Dale Romans	Mr. & Mrs. C. Oliver Iselin III
19	Anneau d'Or	Medaglia d'Oro—Walk Close	32	$453,821	Blaine Wright	Highland Yard LLC
20	Sole Volante	Karakontie—Light Blow	30	$323,310	Patrick Biancone	Flaxman Holdings Limited
21	Finnick the Fierce	Dialed In—Southern Classic	25	$191,290	Rey Hernandez	Paige Jillian Blu Sky Stables
22	Uncle Chuck	Uncle Mo—Forest Music	20	$150,000	Bob Baffert	Stonestreet Thoroughbred Holdings
23	Candy Tycoon	Twirling Candy—Liberty's Lyric	20	$169,850	Todd Pletcher	Jerry Romans Jr.
24	Winning Impression	Paynter—Unbridled Sonya	20	$98,552	Dallas Stewart	WinStar Farm
25	Shotski	Blame—She Cat	19	$236,222	Jeremiah O'Dwyer	Springland Farm & Prime

						Bloodstock LLC
26	South Bend	Algorithms—Sandra's Rose	18	$390,114	Stanley Hough	Highclere, Inc.
27	Necker Island	Hard Spun—Jenny's Rocket	14	$199,730	Stanley Hough	Stonestreet Thoroughbred Holdings LLC
28	Rowdy Yates	Morning Line—Spring Station	7	$346,556	Steven Asmussen	Tracy Rene Strachan
29	Cezanne	Curlin—Achieving	5	$63,000	Bob Baffert	Hill 'n' Dale Equine Holdings, Inc. & St. Elias Stables, LLC
30	Super John	Super Saver—Atlantic Park	2	$72,560	John Servis	Fred W. Hertrich III & John D. Fielding

I found a couple of subjects interesting about this set. One interesting aspect is that three trainers are responsible for two to four contenders each. The last idea is that Uncle Mo was the sire for four out of thirty racers listed. For the individual trainers, I would like to compare their earnings from each horse.

Part 2

Categorical Variables

Now I will look at a few categorical variables in the data. I will create individual frequency tables for the variables.

Let’s start with the Trainers and how many contenders they trained as well as their relative frequency.

Trainer	Frequency	Rel. Freq.
Barclay Tagg	1	3%
Bob Baffert	4	13%
Joe Sharp	1	3%
John Shirreffs	1	3%
Saffie Joseph, Jr.	1	3%
Juan Carlos Avila	1	3%
Todd Pletcher	2	7%
Linda Rice	1	3%

Gustavo Delgado	1	3%
Chad Brown	1	3%
Steven Asmussen	2	7%
Mark Casse	1	3%
Kenny McPeek	1	3%
Ralph Nicks	1	3%
Gregory Foley	1	3%
Peter Eurton	1	3%
Dale Romans	1	3%
Blaine Wright	1	3%
Patrick Biancone	1	3%
Rey Hernandez	1	3%
Dallas Stewart	1	3%
Jeremiah O'Dwyer	1	3%
Stanley Hough	2	7%
John Servis	1	3%
	30	

By looking at the two tables, we can see that there were four individuals that trained more than one horse that qualified for the Kentucky Derby. One interesting aspect is that trainer Bob Baffert trained four of the 30 contenders. He accounts for 13% of the horses alone.

Another categorical variable present in the data, is the breeder. The breeder is the facility credited for the existence of the contender. Looking at the dataset, we can see that there are two facilities that stick out. One being Stonestreet Thoroughbred Holdings LLC and the other is WinStar Farm. Each one was responsible for two out of 30 horses. They exceeded the average per breeder for this data.

The next table provided will be a two-way table. This two-way table will show the corresponse between trainers and breeders.

	Bob Baffert	Todd Pletcher	Steven Asmussen	Stanley Hough	Kenny McPeek	Dallas Stewart	**Total**
Peter E. Blum Thoroughbreds, LLC	1	0	0	0	0	0	1
Hardacre Farm	1	0	0	0	0	0	1
Cloyce C. Clark	0	1	0	0	0	0	1
WinStar Farm	0	0	0	0	1	1	2
Highclere, Inc.	0	0	0	1	0	0	1
Stonestreet Thoroughbred Holdings LLC	1	0	0	1	0	0	2

Tracy Rene Strachan	0	0	1	0	0	0	1
Hill 'n' Dale Equine Holdings, Inc. & St. Elias Stables, LLC	1	0	0	0	0	0	1
Winchell Thoroughbreds	0	0	1	0	0	0	1
Jerry Romans Jr.	0	1	0	0	0	0	1
Total	4	2	2	2	1	1	12

When looking at the table, you can see that the six trainers were related to one of the nine breeders listed on the table. All together they add up to a total of twelve horses listed as contenders. If we look closer, then we can make a drawn-out conclusion. That if we took a horse and had it bred through Stonestreet Thoroughbred Holdings LLC or WinStar Farm, and asked Bob Baffert to train our foal. We could have a possible top 30 contender for the Kentucky Derby.

Part 3

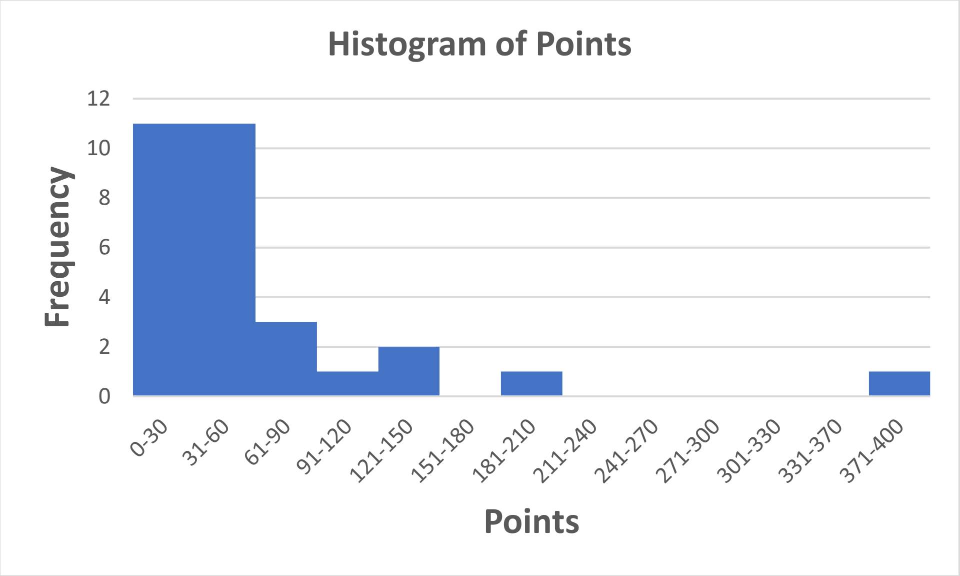
Quantitative Variables

We will start this section by looking at the discrete variables. These variables include the ranking and points of each horse. Ranking corresponds with the points as the horse with most points is ranked first and so on. Points are earned by finishing in the top four in each of the 35 races leading up to the Derby. The top 20 ranked horses get a racing spot in the Kentucky Derby. In the chart provided you will see the summary statistics. Statistics consist of the mean, standard deviation, and the five number summary.

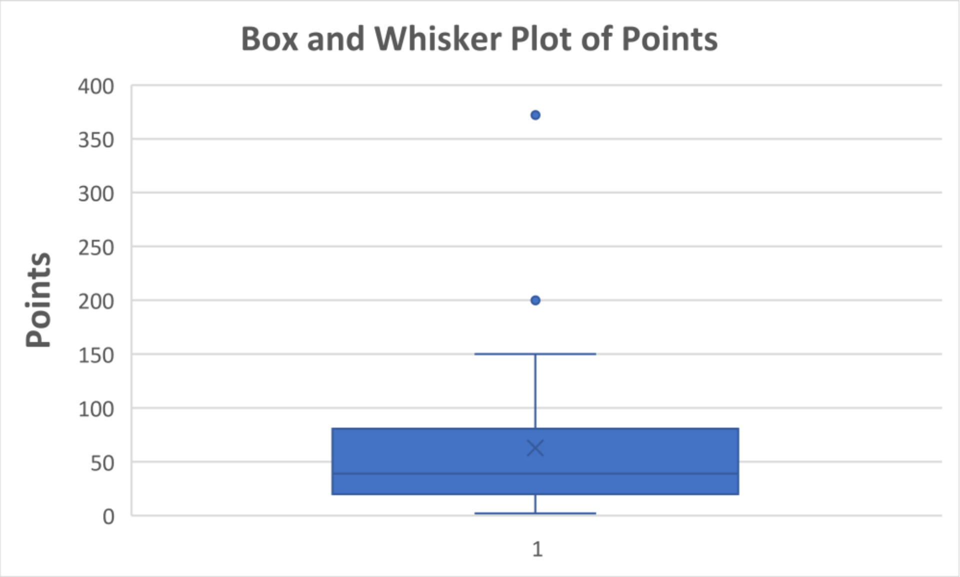
First, we'll determine the summary statistics for Points.

Summary Statistics	Values
Mean:	62.8
Standard Deviation:	74.534
Minimum:	2
Q1:	80
Median:	39
Q3:	20
Maximum:	372

With this chart, we can find the range between contenders and their points. The range of this dataset is 370 points. Now let's make a histogram showing the frequency of points among our data.



The histogram represents a dataset that is skewed. This dataset is positively skewed. I will form a box and whisker plot. This plot shows that we have two outliers within the points data.

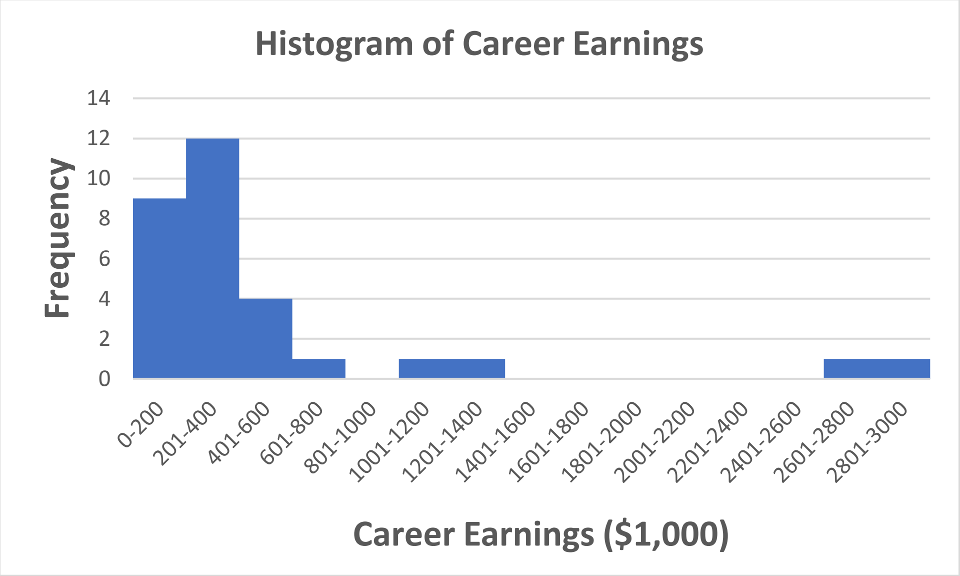


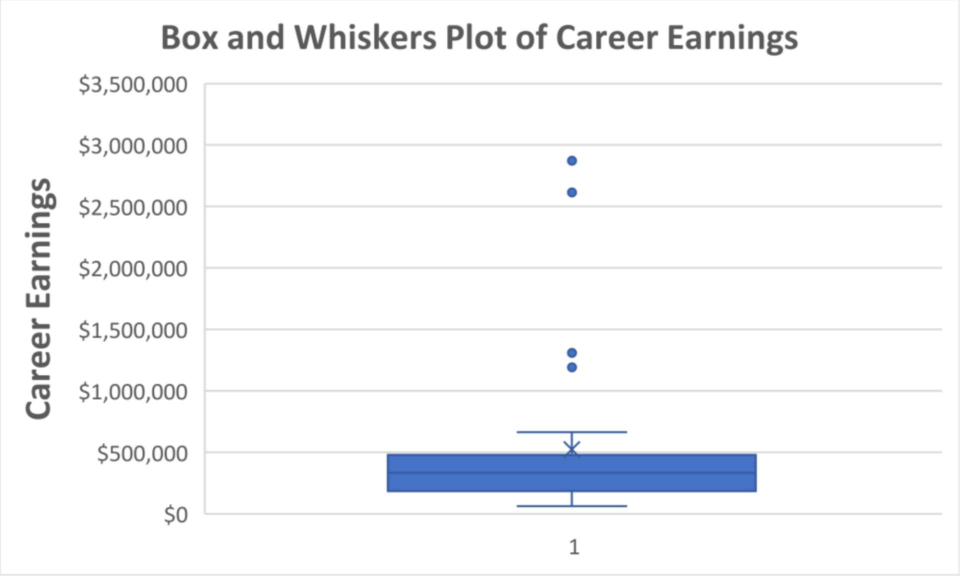
Next, I will look at the Career Earnings in the same way we did the points. Let's see what we get for our summary statistics.

Summary Statistics	Values
Mean:	$524,281
Standard Deviation:	666761.595

Minimum:	$63,000
Q1:	$191,290
Median:	$333,675
Q3:	$463,500
Maximum:	$2,871,200

It's amazing the range of different earnings each horse has acquire throughout their career. The range for this data set is $2,808,200. That's a huge difference. Now I'll make a histogram and a box and whiskers plot for this data.





By studying these two graphs, we can conclude that the data is positively skewed. I can also see that we have four outliers represented by dots outside of the box and whisker plot. But that is not necessarily a bad thing. Those four outliers made the most from their careers on the race track.

Part Four

Hypothesis Testing

For this first hypothesis, I would like to follow with the first interest of this dataset. I want to see where the four trainers stand against the other trains in the top 30 as a whole. I will test this by combining their earnings to calculate the mean and compare it to the mean of the top 30.

$$H_0 : \mu_{FourEarnings} = \$524,281$$

$$H_a : \mu_{FourEarnings} \neq \$524,281$$

For my second hypothesis, I will compare Bob Baffert to what we would expect to see from a trainer in the top 30. I will do this by looking at Baffert’s frequency on the table. Usually you would expect to see a trainer have a frequency of one contender or 1/30.

$$H_0 : P_{BobBaffert} = 1/30$$

$$H_a : P_{BobBaffert} \neq 1/30$$

Now let’s see how these four compares to the rest. Is Bob Baffert really as good as he looks?

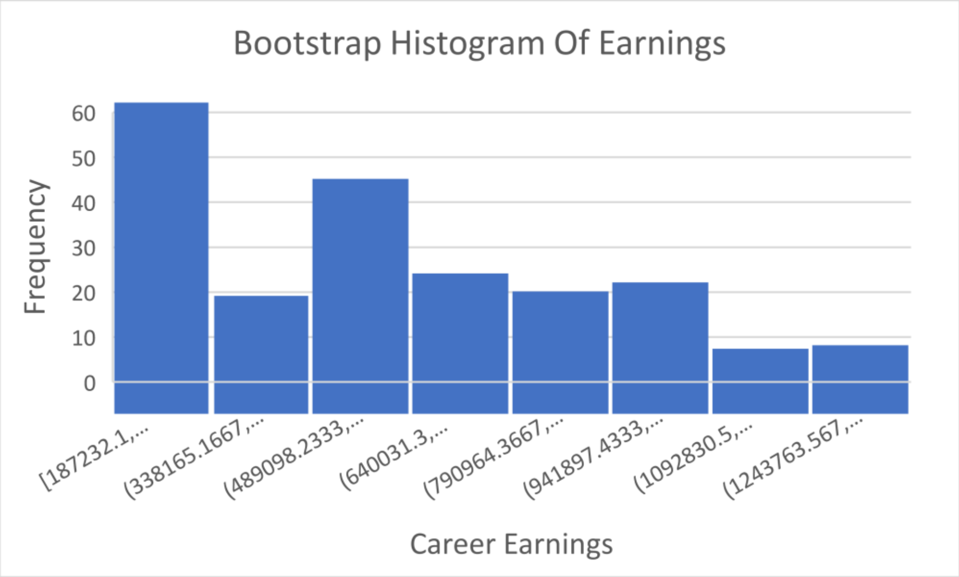
Part Five

Bootstrapping

We will first start with the bootstrap of the quantitative variable hypothesis.

Bootstrap Statistics	
Original Mean:	$514,494.00
Bootstrap Mean:	$502,254.78
Standard Error:	$234,764.06

Provided is a histogram of the bootstrap distribution.



Now lets look at our 95% confidence interval for this variable.

$$\text{CI} = mean \pm z * se$$

Z in this equation is going to be the z-score of 2, and se will be the standard error. Our 95% confidence interval is going to be:

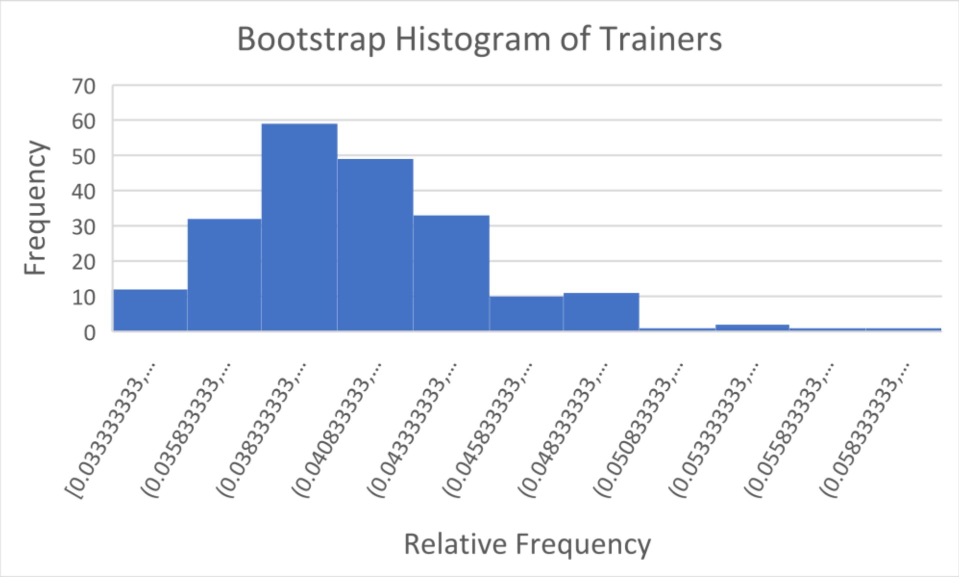
CI = \$524,281 ± \$469,528.12
Or
\$32,726.65 to \$971,782.92

The mean for the overall career earnings is well within the confidence interval. Therefor we can not reject the null hypothesis for this variable.

Next, we will work on the categorical variable hypothesis.

Bootstrap Statistics	
Original Mean:	0.1333
Bootstrap Mean:	0.0415
Standard Error:	0.0043

Here is a histogram of this data.



Now we will calculate the 95% confidence interval with the same equation we used in the previous variable.

$$\text{CI} = 0.1333 \pm 0.0086$$

Or

0.0329 to 0.0502

We would also fail to reject the null hypothesis for our categorical variable.

Part Six

Categorical Hypothesis Test

Let's take a different approach at testing our categorical hypothesis.

$$H_0 : P_{BobBaffert} = 1/30$$

$$H_a : P_{BobBaffert} \neq 1/30$$

To test our hypothesis we will need to find the Z-score. The formula for this is given below.

$$Z = \frac{p\ hat - p}{\sqrt{\frac{p(1-p)}{n}}}$$

We know that *p hat* is the calculated proportion, *p* represents the hypothesized proportion, and lastly *n* is the number that was observed. We should remember the equation for the Standard Error.

$$\mathrm{SE} = \sqrt{\frac{p(1-p)}{n}}$$

Therefore we can interpret the Z-score formula as:

$$\mathrm{Z} = \frac{p\ hat - p}{SE}$$

The Z-score I obtained by using the formulas is 0.00333333. This number is much less than the critical Z-score of 1.96. With this information we can’t reject our null hypothesis and insist that Bob Baffert is present in the top 30 more than other trainers.

Using the formulas provided to get the 95% Confidence Interval, I got a lower value of -0.163 to a upper value of 0.228. Compared to the values we got from our bootstrap, the numbers slightly differ. All together, we get the same conclusion for the 95% confidence interval using both bootstrapping and the Z-score formula.

Part Seven

Quantitative Hypothesis Test

Now we will take the same approach we did in Part Six, but this time we will test our quantitative hypothesis.

$$H_0 : \mu_{FourEarnings} = \$524{,}281$$

$$H_a : \mu_{FourEarnings} \neq \$524{,}281$$

I will use a couple new formulas. First, we will calculate the standard error.

$$SE = \frac{s}{\sqrt{n}}$$

In this formula, the *s* stands for the standard deviation and *n* represents the number observed. The standard deviation for this set is $834,967 and the *n* will be represented as ten. When we plug these values into the formula, we get the SE of $264,040.

Using the new SE, we can construct the 95% Confidence Interval with the following formula.

$$x \pm t * SE$$

With a *t*-value of 1.96, we obtained the following intervals.

$514,494±$ 517517.75
Or
$-3,024.25 to $1,032,011.25

Getting the same conclusion as we had from the bootstrap. The mean is contained within the intervals. Now let’s take a look at our p-value. To get our p-value we need to know our t-value and degree of freedom.

$$t = mean - \mu/SE$$

After using the formula above, I got a t-value of -0.0371 and the degree of freedom is *n-1=9.* Now we plug these two values into the T.DIST function in exel using the cumulative as TRUE. Our p-value is 0.48561987. Comparing this number to alpha= 0.05 for a 95% Confidence Interval. The P-value is much greater than alpha, so we cannot reject our null hypothesis. The same outcome is observed from the bootstrap in Part Six.

Part Eight

Conditional Probabilities

Let’s use our two-way table that was configured in Part Two.

	Bob Baffert	Todd Pletcher	Steven Asmussen	Stanley Hough	Kenny McPeek	Dallas Stewart	**Total**
Peter E. Blum Thoroughbreds, LLC	1	0	0	0	0	0	1
Hardacre Farm	1	0	0	0	0	0	1
Cloyce C. Clark	0	1	0	0	0	0	1
WinStar Farm	0	0	0	0	1	1	2
Highclere, Inc.	0	0	0	1	0	0	1
Stonestreet Thoroughbred Holdings LLC	1	0	0	1	0	0	2
Tracy Rene Strachan	0	0	1	0	0	0	1
Hill 'n' Dale Equine Holdings, Inc. & St. Elias Stables, LLC	1	0	0	0	0	0	1
Winchell Thoroughbreds	0	0	1	0	0	0	1
Jerry Romans Jr.	0	1	0	0	0	0	1

Total	4	2	2	2	1	1	12

Looking at the two-way table, let's find the probability that was bred by a popular breeder. Let's use Stonestreet Thoroughbred Holdings LLC given that Bob Baffert is the trainer. We will use the formula provided below.

$$P(STH\backslash\ Bob\ Baffert) = \frac{P(STH \cap Bob\ Baffert)}{P(Bob\ Baffert)}$$

$$\frac{0.03333333}{0.13333333} = 25\%$$

We can see that these values are dependent and not independent. Trainers are usually given the responsibility of an individual horse before we ever know it will make it in the top 30. Even though it looks to us on the table that Bob Baffert is the best trainer, and Stonestreet Thoroughbred Holdings is the best trainer judging by the number of times they show up in the top 30. We can't determine that they are the best of the best. This second probability is similar to that of the first, but different in little ways. Let's say the trainer is Bob Baffert given that the breeder is Stonestreet Thoroughbred Holdings.

$$P(Bob\ Baffert\backslash STH) = \frac{P(Bob\ Baffert \cap STH)}{P(STH)}$$

$$\frac{0.03333333}{0.06666667} = 50\%$$

Despite what we may conclude from just looking at the data. There is way to determine a winning horse based on just the trainer and lucky breeder. Even though it seems that Stonestreet and Bob Baffert are the best in each category. When they were paired together, their horse ranked 22 out of 30. This horse was just four places away from being able to participate in the Derby. The winning horse depends on multiple standards that the outcome cannot be predicted. Knowing this knowledge is not going to stop individuals from having their fun, but it will however be known that the odds are not always in their favor.