Cumulative Project
Money in Baseball
Class: MDM 4U
Teacher: Mr. Persaud
Student: Eugene Chu
Date: June 5th 2003

Table of Contents

1. Introduction
2. Mind map
3. Salary Data
4. Wins Data
5. Losses Data
6. Single variable graphs: Total Team Salaries
7. Total Team wins
8. Total Team losses
9. Single variable analysis
10. Double variable graphs: Avg. Salary Vs Wins
11. Years Vs. Salary Rank
12. Avg. Salary Vs. Losses
13. Double Variable analysis
14. Probability Distribution: Winning World Series (Salary)
15. Winning World Series (Salary Rank)
16. Making Playoffs (Salary)
17. Making Playoffs (Salary Rank)
18. Winning .500 games
19. Simulation: Winning World Series (Salary), Winning World Series (Salary Rank)
20. Making Playoffs (Salary), Winning .500 games
22. Simulation Results and Expected Value
23. Normal Distribution
24. Cause and Effect Diagram
25. Critical Path analysis
26. Conclusions, possible errors
27. Bibliography

Introduction

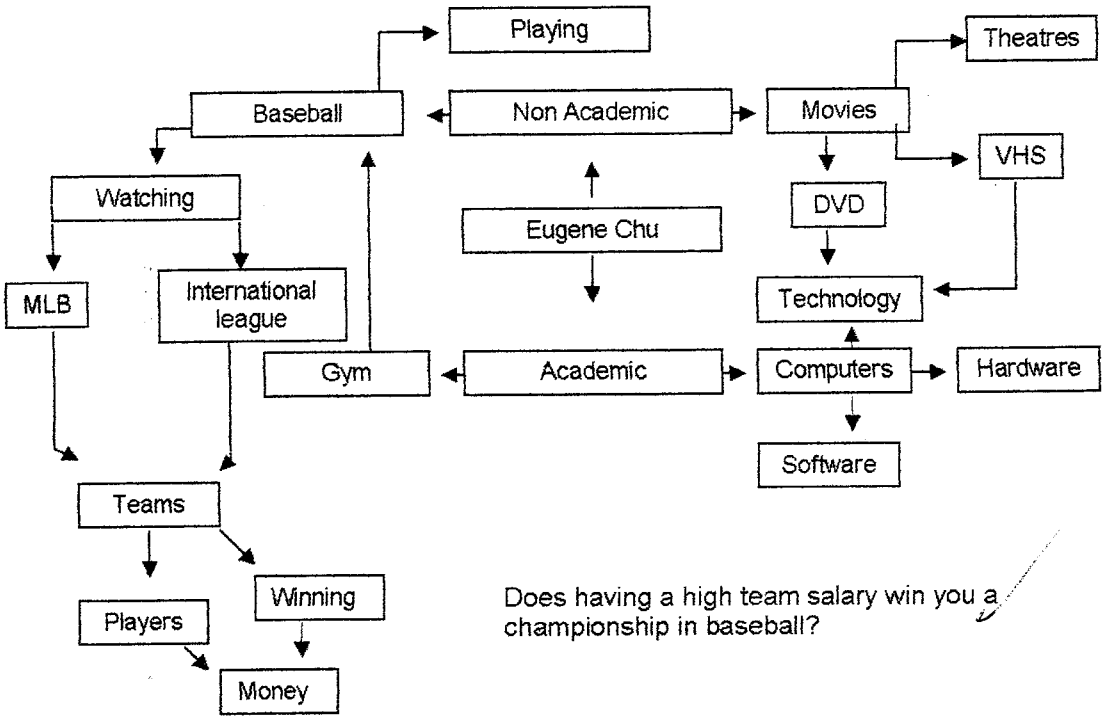
Approximately two billion twenty-four million six hundred thousand dollars later, the MLB 2002 season is finally over. The result of all this money being spent, one team of approximately 30 players celebrating while the rest go home and wait for next year. This is the way of life in most sports and the only thing players have to compensate for their loss is their salary which if you put into perspective quite a nice compensation.

Baseball, the great American past – time, has always had a vast and in depth history that began in the 1800s and has continued strongly up to this day. The birth of Major league baseball (MLB) only fuelled this history, allowing fans of the game to watch their favourite players rack up statistics game after game. It is here in baseball where a simple game of hitting and running has now resulted in billions of dollars being spent each year for star athletes.

Through the exploration of my personal interests, I found several different areas of possible topics to research. These areas included movies, technology and sports. After researching through these topics I found that sports had the broadest and most accessible data. After much thought on the different aspects of sports I narrowed my topic to one specific area, money in sports. Money in sports has always been a largely debated item, but rather than debate whether athletes should be earning the salaries that they receive, I decided to research the reasons why they get these salaries. Ultimately, it became evident that the one goal of an athlete and an organization is to win the championship. From here, I decided which sport I was going to examine, the final result was baseball. Baseball is unique when comparing salaries with other sports. This is mainly because baseball does not have a salary cap, the more money you make the more money you can spend. This ability allows owners to purchase players for their teams without restrictions leading to possible unfair teams. These "unfair" teams have lead to recent trends, for example the New York Yankees have won the World Series on several occasions in the past few years and "coincidently" spent the most on players for those years. Rather than looking at individual salaries I chose to look at team salaries because individuals don't win championships, teams win championships.

The purpose of my project is to discover trends and relationships between winning a championship and team salaries. Ultimately, I would like to prove that yes, spending larger amounts of money results in a championship which would answer the following question. Does having a higher team salary result in more wins, and ultimately a championship?

Mind Map



Does having a high team salary win you a championship in baseball?

Salary Data

	Team Salary Per Year												
Team	2002	2001	2000	1999	1998	1997	1996	1995	1994	1993	1992	Mean	Median
ANGELS	61721667	47735168	55766667	49893166	38537000	29452672	26892500	28974167	20691500	27230334	33529854	38220427	35875140
DIAMONDBACKS	102819999	85247999	77880333	70370999	29161500	n/a	n/a	n/a	n/a	n/a	n/a	73096166	75488250
BRAVES	93470367	91936166	82732500	75065000	59536000	50488500	47930000	45199000	40502167	38131000	32975333	59815094	55012250
ORIOLES	60493487	74279540	83141198	70818363	70408134	54871399	48726832	40835519	37669769	26914000	20997667	53559628	54215514
RED SOX	108366060	109675833	81210333	71720000	51647000	43232000	39676000	28672250	36334084	37108583	42203584	59076884	47439500
CUBS	75690833	64515833	62129333	55368500	49383000	39829333	30954000	32460834	35717333	38303166	29060833	46673909	43251621
WHITE SOX	57052833	65628667	31159000	24550000	36840000	54377500	41940000	39632834	38416836	34598166	28413500	41146303	39024835
REDS	45050390	48784000	44217500	42142761	21995000	46267000	40719334	37240667	39826333	42851167	35203999	40390741	41431048
INDIANS	78909449	92660001	76508334	73857962	59033499	54130232	45317914	35185500	28490167	15717667	8236166	51640626	52885429
ROCKIES	56851043	71541334	64130857	54392504	47433333	42870501	34918490	31146135	22979000	8829000	n/a	43509220	43509220
TIGERS	55048000	49356167	61740167	34959666	22625000	16304500	21941000	35862501	40042501	36548166	28773834	36654682	36205334
MARLINS	41979917	35562500	19870000	15150000	33434000	47753000	30079500	23670000	20275500	18106545	n/a	28588096	28588096
ASTROS	63448417	60387667	52356667	55289000	40629000	32935000	26894000	31624000	32041500	28854500	13352000	39801068	36368034
ROYALS	47257000	35422500	23132500	16527000	32912500	31225000	18480750	27608834	40481334	40102666	33643834	31526720	32219610
DODGERS	94850953	109105953	90975953	71135786	47970000	43400000	34647000	30459001	37194001	37833000	43788166	58305438	45879083
BREWERS	50287833	45099333	35782833	42927395	32252583	23320332	20482000	16189600	23375513	22948834	30253668	31174539	30714103
TWINS	40225000	24130000	15654500	16355000	26182500	25747500	21961500	24527500	27641500	27284933	27432834	25194797	25471149
EXPOS	38670500	34849500	33527666	16363000	9202000	18335500	15410500	12031000	18955000	14881334	15869667	20735970	17349250
METS	94633593	93674428	79759762	71331425	49559665	38474567	23456500	24301440	29890324	38350167	44352002	53434898	46955834
YANKEES	125928583	112287143	92938260	88130709	63159898	59148877	52189370	46657016	44785334	41305000	35966834	69317911	61154388
ATHLETICS	40004167	33810750	32121833	24150333	20063000	21911000	19404500	35961500	33169500	35565834	39957834	30556386	32645667
PHILLIES	57954999	41663833	46947667	30516500	36085000	35463500	28393500	28580000	31422000	26812334	23804834	35240379	33331189
PIRATES	42323599	57760833	29561667	24217666	13752000	9071666	21253500	17043000	20265500	23565667	32589167	26491297	23891667
PADRES	41425000	38882833	54971000	45932179	45368000	34698672	27133026	25923334	13529333	24557333	27584167	34545898	34622285
GIANTS	78299835	63280167	53541000	46059557	40320835	33469213	34605225	34931849	40054300	34567500	33126168	44750514	40187568
MARINERS	80282668	74720834	59215000	44371336	52032291	39667628	39221501	34241533	27872167	31616333	22483834	45975011	42019482
CARDINALS	74660875	78333333	63993023	46248195	52572500	44179167	38741666	30956000	28956001	22615334	26889836	46195085	45187126
DEVIL RAYS	34380000	56980000	64407910	37812500	25317500	n/a	n/a	n/a	n/a	n/a	n/a	43779582	40796041
RANGERS	105726122	88633500	70785000	81301598	54704595	50112268	35862028	32367226	32423097	35641959	29740667	56118005	52408432
BLUE JAYS	76864333	76895999	46363332	48165333	48415000	45894833	28486708	49791500	41937668	45747666	43663666	50202367	47264333
Standard Deviation of Mean	12921347												

The table above is the data that I have collected concerning team salaries. It shows each team's salaries for the past 10 years, as well I have calculated the mean (=average(X: Y)) and median (=median(X: Y)) for each team. As a result of the data being continuous since the numbers are so large and have a large range, I have decided not to do mode because none of the numbers are reoccurring. Also, I have calculated the standard deviation based on the mean using the formula =stdeva(X: Y). X and Y represent the cells that were used to calculate the values (mean, median and standard devation).

Wins Data

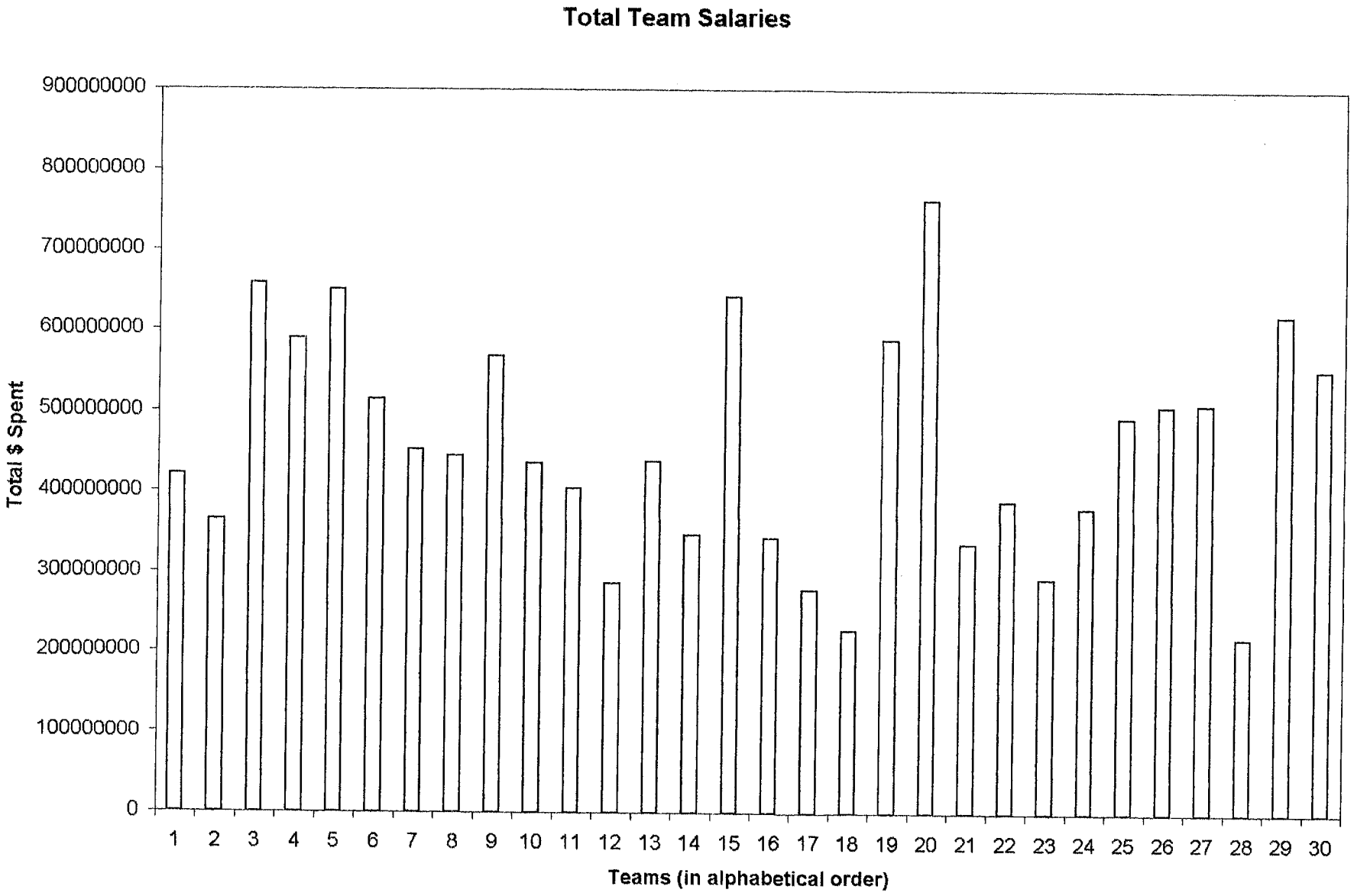
Teams	Wins per year														
	2002	2001	2000	1999	1998	1997	1996	1995	1994	1993	1992	Total Wins	Mean	Median	Mode
Angels	99	75	82	70	85	84	70	78	47	71	72	833	76	75	70
Diamondbacks	98	91	85	100	65	n/a	n/a	n/a	n/a	n/a	n/a	439	88	91	#N/A
Braves	101	88	95	103	106	101	96	90	68	104	98	1050	95	98	101
Orioles	67	63	74	78	79	98	88	71	63	85	89	855	78	78	63
Red Sox	93	82	85	94	92	78	85	86	54	80	73	902	82	85	85
Cubs	67	88	65	67	90	68	76	73	49	84	78	805	73	73	67
White Sox	81	83	95	75	80	80	85	68	67	94	86	894	81	81	80
Reds	78	66	85	96	77	76	81	85	66	73	90	873	79	78	66
Indians	74	91	90	97	89	86	99	100	66	76	76	944	86	89	76
Rockies	73	73	82	72	77	83	83	77	53	67	n/a	740	74	75	73
Tigers	55	66	79	69	65	79	53	60	53	85	75	739	67	66	79
Marlins	79	76	79	64	54	92	80	67	51	64	n/a	706	71	71.5	79
Astros	84	93	72	97	102	84	82	76	66	85	81	922	84	84	84
Royals	62	65	77	64	72	67	75	70	64	84	72	772	70	70	64
Dodgers	92	87	86	77	83	88	90	78	58	81	63	883	80	83	#N/A
Brewers	56	68	73	74	74	78	82	65	53	69	92	784	71	73	74
Twins	94	85	69	63	70	68	78	56	53	71	90	797	72	70	#N/A
Expos	83	68	67	68	65	78	88	66	74	94	87	838	76	74	68
Mets	75	82	94	97	88	88	71	69	55	59	72	850	77	75	88
Yankees	103	95	87	98	114	96	92	79	70	88	76	998	91	92	#N/A
Athletics	103	102	91	87	74	65	78	67	51	68	96	882	80	78	#N/A
Phillies	80	86	65	77	75	68	67	69	54	97	70	808	73	70	#N/A
Pirates	72	62	69	78	69	79	73	58	53	75	96	784	71	72	69
Padres	66	80	76	74	98	76	91	70	47	61	82	821	75	76	76
Giants	95	89	97	86	89	90	68	67	55	103	72	911	83	89	89
Mariners	93	116	91	79	76	90	85	79	49	82	64	904	82	82	79
Cardinals	97	93	95	75	83	73	88	62	53	87	83	889	81	83	83
Devil Rays	55	62	69	69	63	n/a	n/a	n/a	n/a	n/a	n/a	318	64	63	69
Rangers	72	73	71	95	88	77	90	74	52	86	77	855	78	77	77
Blue Jays	78	80	53	84	88	76	74	56	55	95	96	835	76	78	#N/A
Standard Devation based on mean	6.9237														

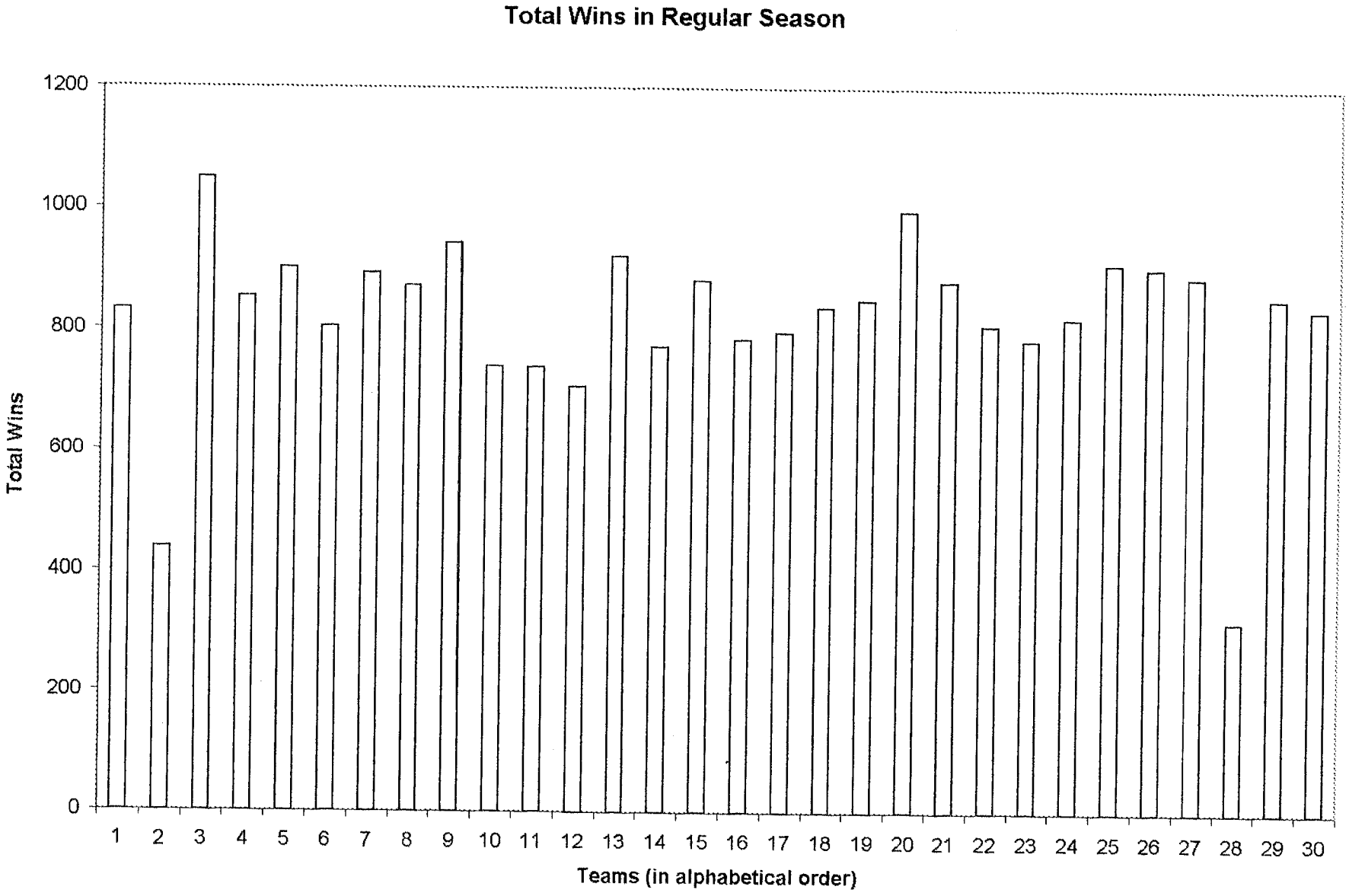
This chart displays data concerning wins of a team. It shows the year by year wins for each team. Like the previous chart it has mean (=average(X: Y)) and median (=median(X: Y)), but because this data is discrete, since a team can only win a maximum of 182 games making the range of wins 0-182, I've decided to calculate the mode (=mode(X: Y)). Although, in some cases a mode was not found.

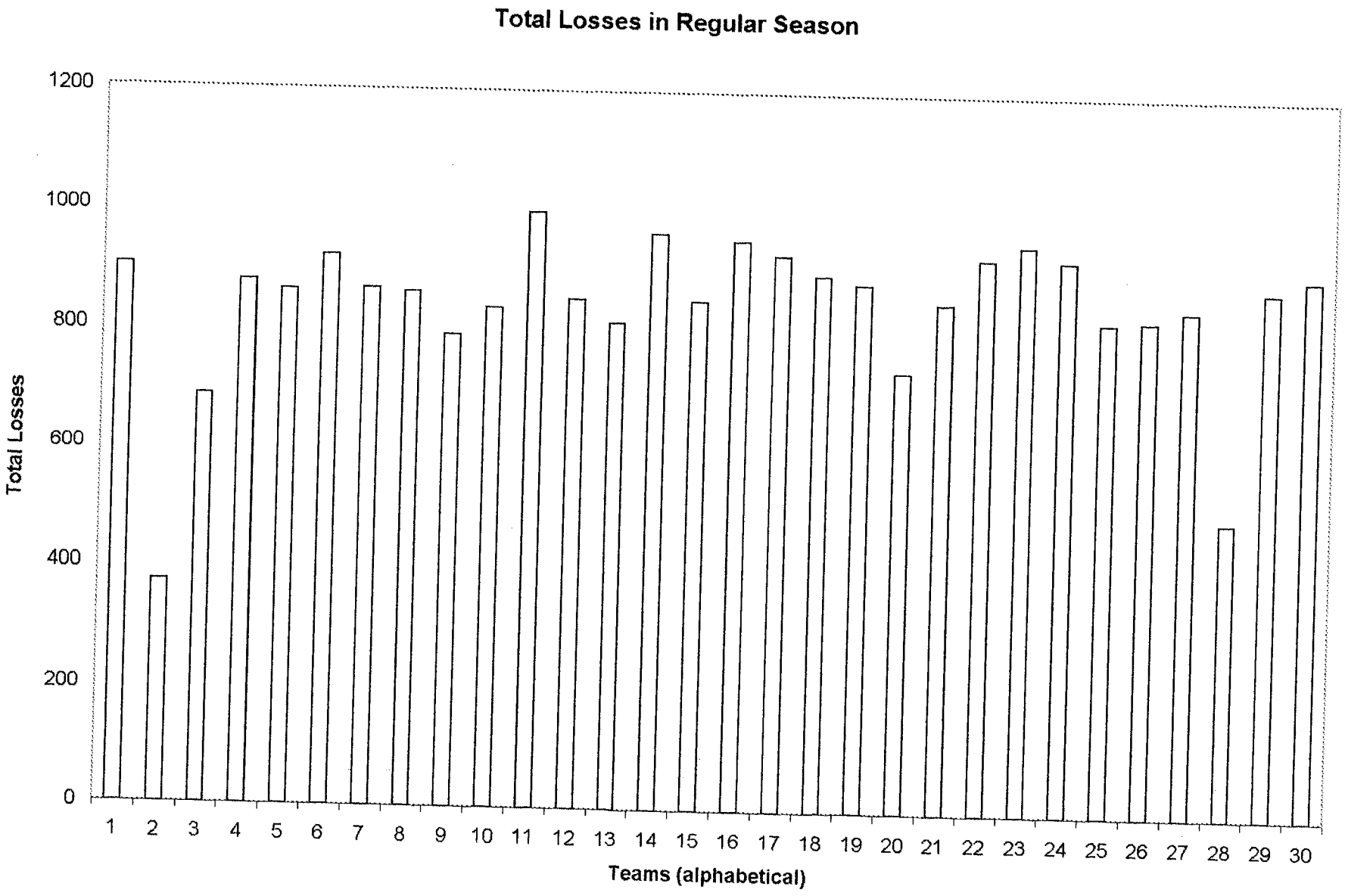
Lo. Data

	Losses per year														
Teams	2002	2001	2000	1999	1998	1997	1996	1995	1994	1993	1992	Total losses	Mean	Median	Mode
Angels	63	87	80	92	77	78	92	84	68	91	90	902	82	84	92
Diamondbacks	64	71	77	62	97	n/a	n/a	n/a				371	74	71	#N/A
Braves	61	74	67	59	56	61	66	72	46	58	64	684	62	61	61
Orioles	95	99	88	84	83	64	74	91	49	77	73	877	80	83	#N/A
Red Sox	69	80	77	68	100	84	77	76	61	82	89	863	78	77	77
Cubs	95	74	97	95	72	94	86	89	64	78	76	920	84	86	95
White Sox	81	79	67	87	102	82	77	94	46	68	84	867	79	81	#N/A
Reds	84	96	77	66	85	86	81	77	48	89	72	861	78	81	77
Indians	88	71	72	65	73	76	63	62	47	86	86	789	72	72	86
Rockies	89	89	80	90	85	79	79	85	64	95		835	84	85	89
Tigers	107	96	83	93	97	83	109	102	62	77	87	996	91	93	83
Marlins	83	86	83	98	108	70	82	95	49	98		852	85	84.5	83
Astros	78	69	90	65	60	78	80	86	49	77	81	813	74	78	78
Royals	100	97	85	98	90	95	87	92	51	78	90	963	88	90	90
Dodgers	70	75	76	85	79	74	72	84	56	81	99	851	77	76	#N/A
Brewers	106	94	89	88	88	84	80	97	62	93	70	951	86	88	88
Twins	68	77	93	99	92	94	84	106	53	91	72	929	84	91	#N/A
Expos	79	94	95	94	97	84	74	96	40	68	75	896	81	84	94
Mets	87	80	68	65	74	74	91	93	58	103	90	883	80	80	74
Yankees	59	67	75	64	48	66	70	83	43	74	86	735	67	67	#N/A
Athletics	59	60	71	75	88	97	84	95	63	94	66	852	77	75	#N/A
Phillies	82	76	97	85	87	94	95	93	61	65	92	927	84	87	#N/A
Pirates	90	100	93	84	93	83	89	104	61	87	66	950	86	89	93
Padres	96	82	86	88	64	86	71	92	70	101	90	926	84	86	86
Giants	67	73	65	76	73	72	94	95	60	59	90	824	75	73	73
Mariners	69	46	71	83	86	72	77	83	63	80	98	828	75	77	83
Cardinals	65	69	67	87	79	89	74	100	61	75	79	845	77	75	79
Devil Rays	107	100	93	93	99	n/a	n/a	n/a				492	98	99	93
Rangers	90	89	91	67	74	85	72	88	62	76	85	879	80	85	85
Blue Jays	84	82	109	78	74	86	88	106	60	67	66	900	82	82	#N/A
Standard Devation based on mean	6.98														

The data above illustrates the losses suffered by each team for the past 10 years. Like the wins data, this is discrete because a team can only lose 162 games, making a range of 0-162. This allows me to calculate mean (=average(X: Y)), median (=median(X: Y)) and mode (=mode(X: Y)). Note: Some teams were non-existent during the 10 years, as well 162 games were not played in 1994, neither was a World Series due to a strike.


Total Team Salaries
900000000
800000000
700000000
600000000
500000000
400000000
300000000
200000000
100000000
0
Total $ Spent
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
Teams (in alphabetical order)


Total Wins in Regular Season
1200
1000
800
600
400
200
0
Total Wins
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
Teams (in alphabetical order)


Total Losses in Regular Season
1200
1000
800
600
400
200
0
Total Losses
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
Teams (alphabetical)

Single Variable Analysis

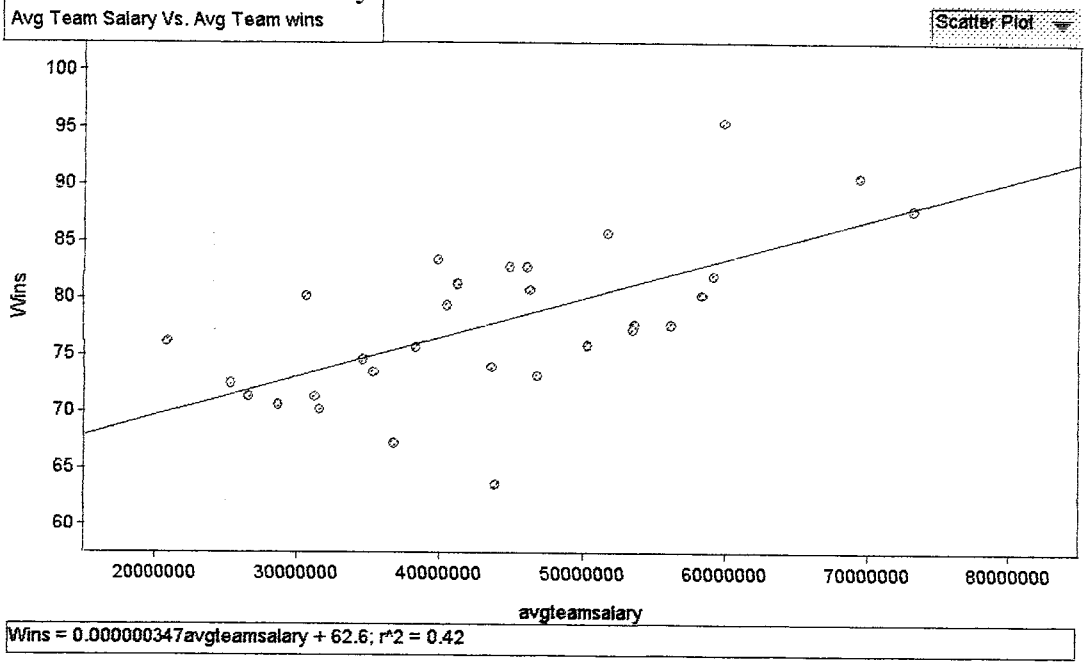
The following three graphs were simply single variable graphs that the total salaries, wins and losses for each team for the past 10 years. The teams were listed in alphabetical order based on their city names and the corresponding total salaries were placed appropriately. The first graph, total team salaries, displays continuous data since the range are so vast. This data was found from a secondary source found on the net (http://asp.usatoday.com/sports/baseball/salaries/default.aspx). The second graph, total team wins is discrete data because there is a limit to how many wins a team can have. Again, this was found using a secondary source found on the net (http://baseball1.com/c-stats.html). Finally the last graph, total team losses, is also discrete since a team can only have a certain amount of losses. Like the previous two graphs this was also found using a secondary source on the net (http://baseball1.com/c-stats.html).

Analyzing the graphs was fairly simple; I looked at the largest bars and found which teams had the highest values. For the first graph team 20 (Yankees) have spent the largest amount of money on players, while team 28(Devil Rays) have spent the least. Although, you should note that the Devil Rays have not been a franchise for the full 10 years. Other than a few exceptions the rest of the graph was rather uniform. For the second graph, team 3 (braves) have the most wins, while team 28 (Devil Rays) had the least. Again they are last due to the fact that they have not been a team until recently. Finally the third graph shows that team 11(tigers) have the most losses and that team 28 (Devil) rays have the least losses. The last two graphs are similar to the first one, in that they both have a few outliers, but other than that, they depict a uniform graph.

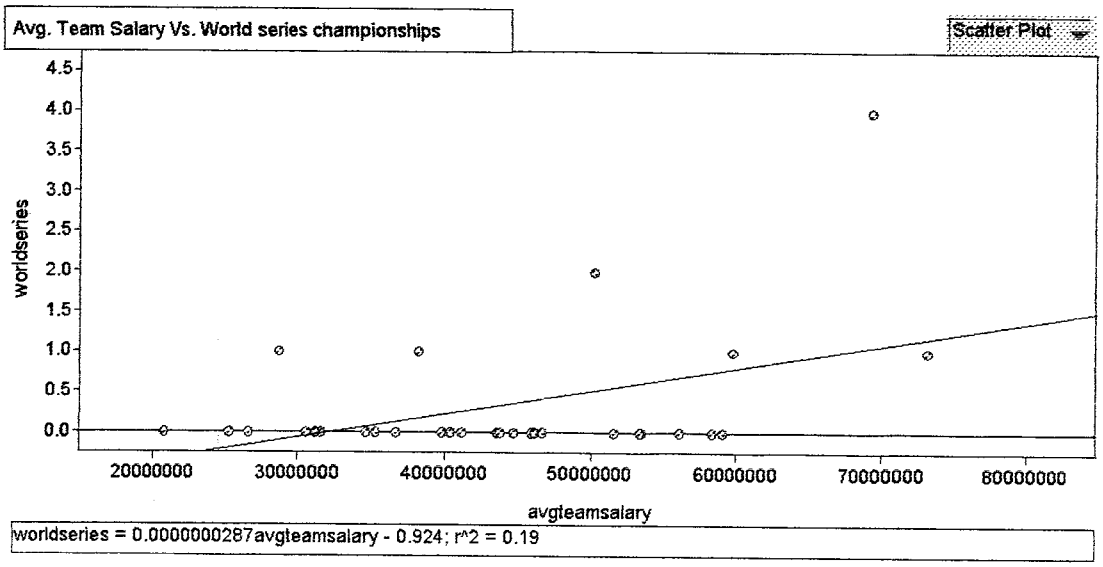
The previous three graphs were not analyzed in depth since they don't really show too much. All actual data can be obtained from the charts provided earlier. Every graph displays qualitative data since they are all numbers. These single variable graphs must be compared with one another in order to find relationships, correlations and trends. In order to find these items, double variable scatter plots must be used; the next section will cover these topics.

Eugene Cha

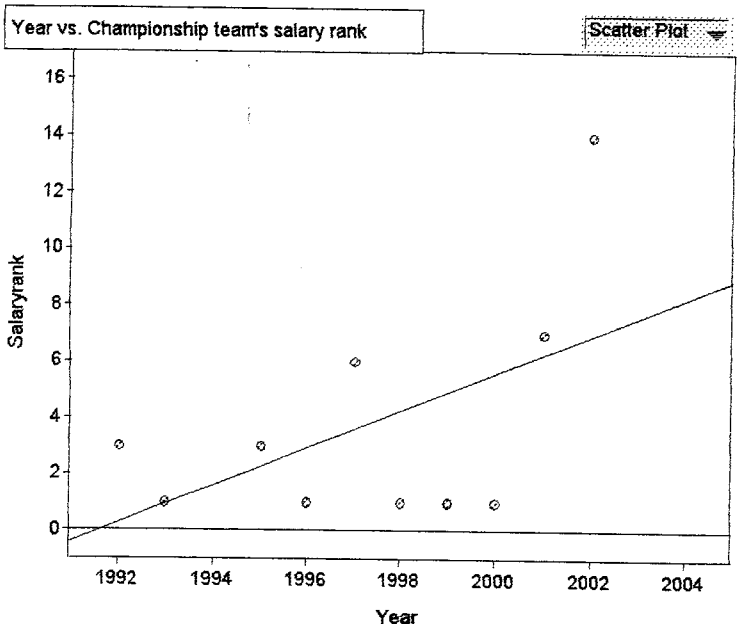
Double Variable Analysis



The scatter plot above shows that there is a linear relationship between the avg. team salary and avg. team wins. Teams with lower team salaries tend to have less wins in the regular season, while the majority of teams that spend more money on their players have more wins. Although the graph portrays an obvious positive correlation between the variables, it is important to mention that the r^2 which means how well the data fits the line of best fit, is only 0.42. This means that the correlation between the variables is not that strong, since the strongest correlation has a r^2 of 1. Also, it's interesting to note that the team with the lowest avg team salary does not have the lowest amount of avg wins, infact it has quite a large amount of wins. The lowest amount of wins belongs to a team that has an average salary that is roughly in the middle. Finally, the team that has the most amount of wins in the regular season does not have the highest team salary, this team has spent only slightly more than most of the other teams. Most teams have spent within 40- 60 million dollars, the atlanta braves have spent roughly 60 million and have the most amount of wins.



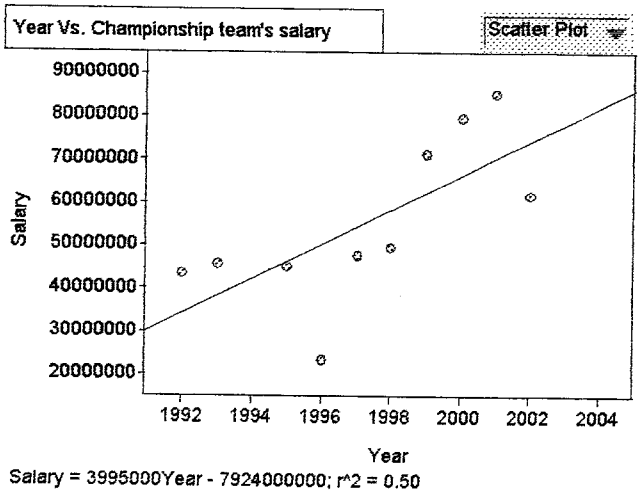
The scatter plot above shows the amount of world series championships vs. the avg team salary. This graph is vital to the project because the ultimate goal for any team is to win a championship, it's far more important than getting the most amount of regular season wins. Due to the fact that the graph only displays data from the past 10 years, (1992- 2002) many of the teams only 0 championships. Some important points to note are that the top 3 teams with the highest average salaries have one at least 1 championship, with the new york yankees winning 4. As well one team with an extremely low avg team salary was able to win a championship and the other championships were given to teams that spent roughly the same amount of money as the other teams. Also the r^2 is very low meaning that the correlation is very weak, but that is most likely due to the fact that many of the teams have not won any championships.



Salaryrank = 0.6653Year - 1325.1; r^2 = 0.28

The graph to the left shows the rank of the champion team's salary within the league. This graph is important because the fluctuation amoung team salaries is very large and this graph helps to show the amount of dollars spent for the particular year in which that team won the world series. For example, during 1998, 1999 and 2000, the New york yankees won the world series and were ranked 1st with the highest team salary for those years. The graph allows individuals to see where each championship team stood among salary. Some interesting things to note are that from 1992-1996, the championship team was always within the top 3 in salary rank. Also, besides the yankees from 1998-2000, every championship team afterwards has had a low ranking. Finally, the 2002 champions, anaheim angels had a ranking of 14 which means 13 teams spent more money on players then them but they were still able to win the championship.

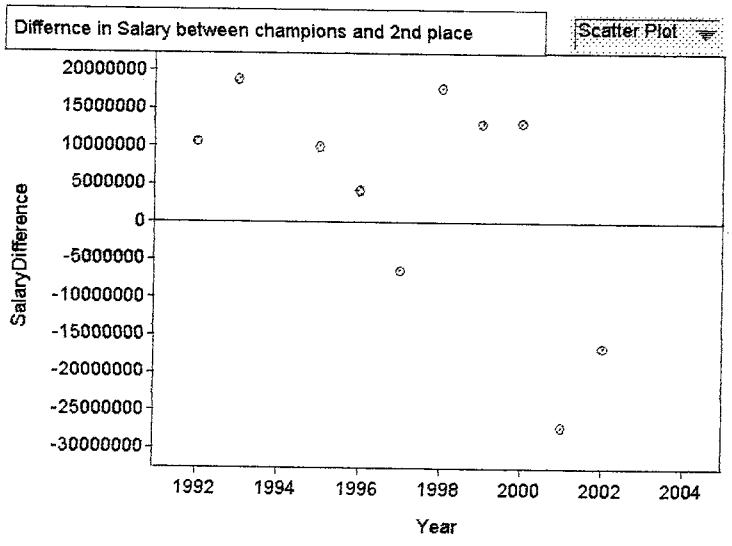
The scatter plot to the left is a continuation of the graph above. It simply states the actual amount of dollars spent by the championship team for that particular year. While it's not directly related to the topic, it allows individuals to see the amounts instead of rank so that they can get an idea of how much was spent that year. An interesting thing to note is that the championship team's salary seems to be increasing year after year, giving a message that in the coming years, teams will have to spend more money inorder to win the world series.



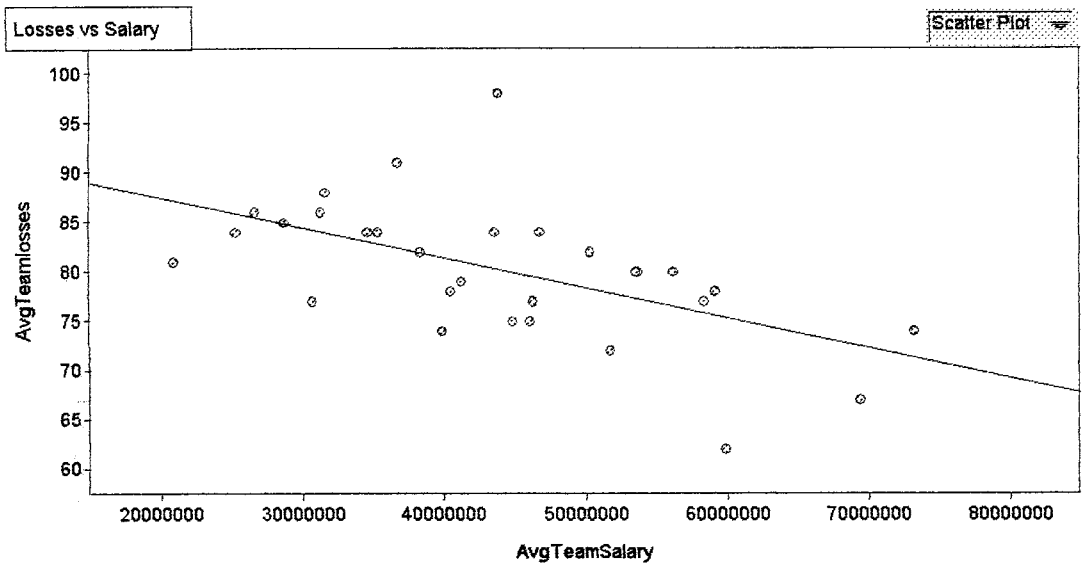
Salary = 3995000Year - 7924000000; r^2 = 0.50

Differnce in Salary between champions and 2nd place

	Year	SalaryDiff...	<new>
1	1992	10688333	
2	1993	18935332	
3	1994		
4	1995	10013500	
5	1996	4259370	
6	1997	-6377232	
7	1998	17791898	
8	1999	13065709	
9	2000	13178498	
10	2001	-27039144	
11	2002	-16578168	

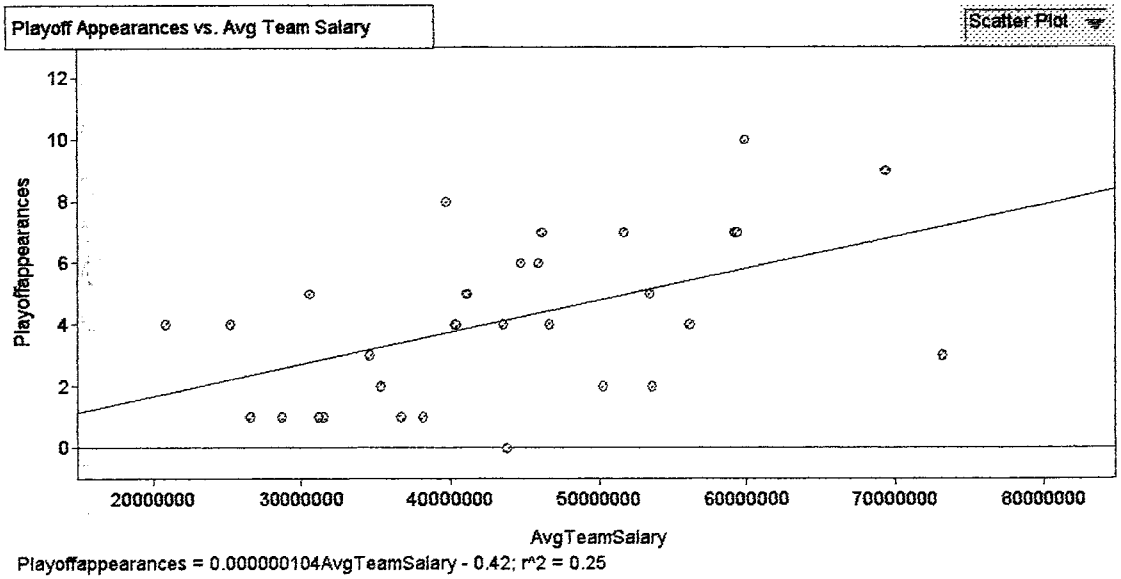


The graph and table above show the difference in salary of the championship team and the runner up. I took the championship team's team salary for the year they won and subtracted the 2nd place team's salary from it. The graph helps to show the range between the 1st and place and 2nd place teams. 7 out of 10 years the championship team had a larger team salary ranging from 5 - 20 million dollars more. But there were times when the 2nd place team had spent more money than the 1st place team. These are indicated by the negatives values on the graph. The values range from 5 - 30 million dollars. This graph shows that there have been times when the world series champion was not determined by the amount of money they had spent.



AvgTeamlosses = -0.000000304AvgTeamSalary + 93; r^2 = 0.32

The scatter plot above displays the correlation between avg. team salary and team losses. As expected the correlation is negative, with teams who spend more on players having fewer losses throughout the season. This graph displays a weak negative correlation as the equation of the line shows that the slope is fairly small. As well the r^2 value of 0.32 tells you that the line of best fit does not represent the points very well, leaving a large chance of error. It's interesting to note that most of the points are roughly uniform and should be able to be classified as a normal distribution. Also, the lowest amount of losses belongs to a team that has spent slightly more than a large portion of the other teams. This team (atlanta braves) is proves that there is a chance of winning alot of games and losing few while still only spending roughly the same as most other teams. Another thing to note about my last comment is that the team that has the lowest losses has the third highest avg. team salary, although just slightly.



Playoffappearances = 0.000000104AvgTeamSalary - 0.42; r^2 = 0.25

In all sports, teams who are successful in the regular season are not rewarded a championship directly. Instead they earn a shot at the championships through the playoffs. The following graph compares a team's salary and the number of times that team has made it into the playoffs. There is clearly a positive correlation that leads us to believe that teams who spend more, get into the playoffs more. One thing to note is that the team with the highest mean salary has only 3 playoffs appearance, but you should also note that this team is the Arizona Diamondback, and they have only been a franchise for 4 years, yet they have made the playoffs 3 of the last 4 years. Also, most expansion teams don't make it into the playoffs until after several years, but the Diamondbacks made it into the playoffs in their second year and the only time they didn't make the playoffs was when they had a very low team salary.

The previous graphs were used to compare variables, find relationships and trends. All variables were quantitative and all were discrete, except the average team salary, which was continuous. For all the graphs, I used the charts provided at the beginning of the project and manipulated the data using Fathom. I used the drag and drop function to produce the scatter plots and then I used the least-square lines feature to plot the line of best fit and find the equation of the line and r^2 value. Most of the graphs showed positive correlations with only one showing a negative correlation. As well, all the graphs had weak relationships and outliers that made the line of best fit inaccurate a times.

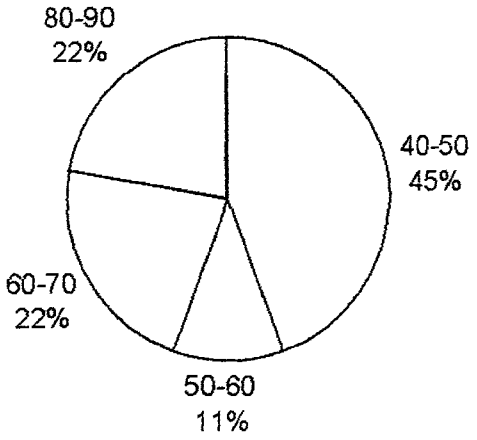
I've already discussed some of the interesting and important points that the graph has pointed out. I will just review a couple that I feel are vital to answering my thesis question. Firstly, the Avg. Salary vs. Avg. team wins graph shows that there is a correlations between salary and regular season wins. As well, besides a few outliers, most of the data follows this trend, so it may be safe to say that higher team salaries result in more wins. Secondly, the second graph shows that there is a small correlation between team salary and World Series championships. The only problem is that this correlation is very weak and if you exclude the Yankees, there really isn't much of a correlation. While I can not come to conclusions at this point, it seems that team salaries doesn't have a major effect on winning major league championships, since the main reason the graph appears the way it does is because of one outlier, the New York Yankees. Finally the last graph comparing Avg. Team Salaries vs. Playoff Appearances displays a positive correlation that is probably the strongest compared to all the others. It seems as though spending a lot of money does get a team into the playoffs. Even an expansion team who had a high team salary was able to make it into the playoffs and newly formed teams rarely ever have success their first few seasons.

Coming to conclusions at this point is a bit premature as a lot of the data is pointing towards my thesis that yes, higher team salaries results in more success, but there is also exceptions in these graphs that prove otherwise. Moving on to the probability distributions and simulations may help clarify the answer to my question.

Probability Distribution

Team Salary (in millions)	# of World Series Championships	Probability of winning a world series
10-20	0	0%
20-30	0	0%
30-40	0	0%
40-50	3	30%
50-60	1	10%
60-70	2	20%
70-80	0	0%
80-90	2	20%
90-100	1	10%
100+	0	0%

Probability of winning a World Series

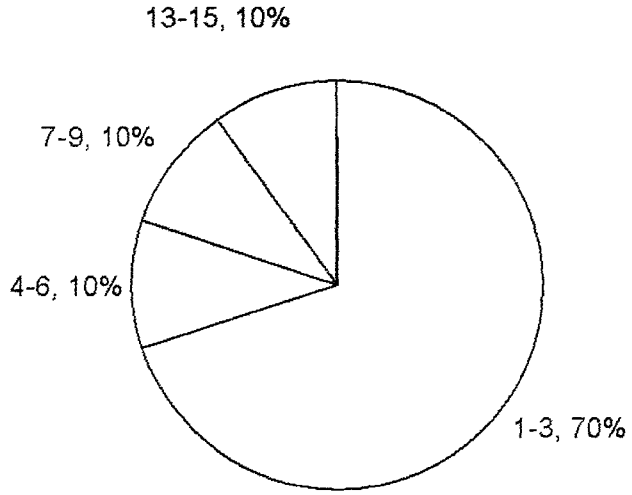


The table above shows the probability of winning a world series based on team salary for the past 10 years. In order to complete the table I looked at the salary of the championship team for each of the 10 years and placed them into their appropriate salary range. For example when the blue jays won in 1992 they spent 43,663,666 so I placed their victory into the 40-50 million dollar range. From this pie graph, it seems as though spending 40-50 million dollars results in a greater chance at winning a world series. The problem with this graph is that it does not factor in the fact that the average team salary for that particular year might have been low. For example the average amount spent in 1992 was 30,149,767, but the average amount spent in 2002 was 67,489,251. This shows that it's possible that 40-50 million may have been a lot for the years that those teams won and that spending 40-50 million for the 2004 season will not necessarily result in a 45% chance of winning the championship.

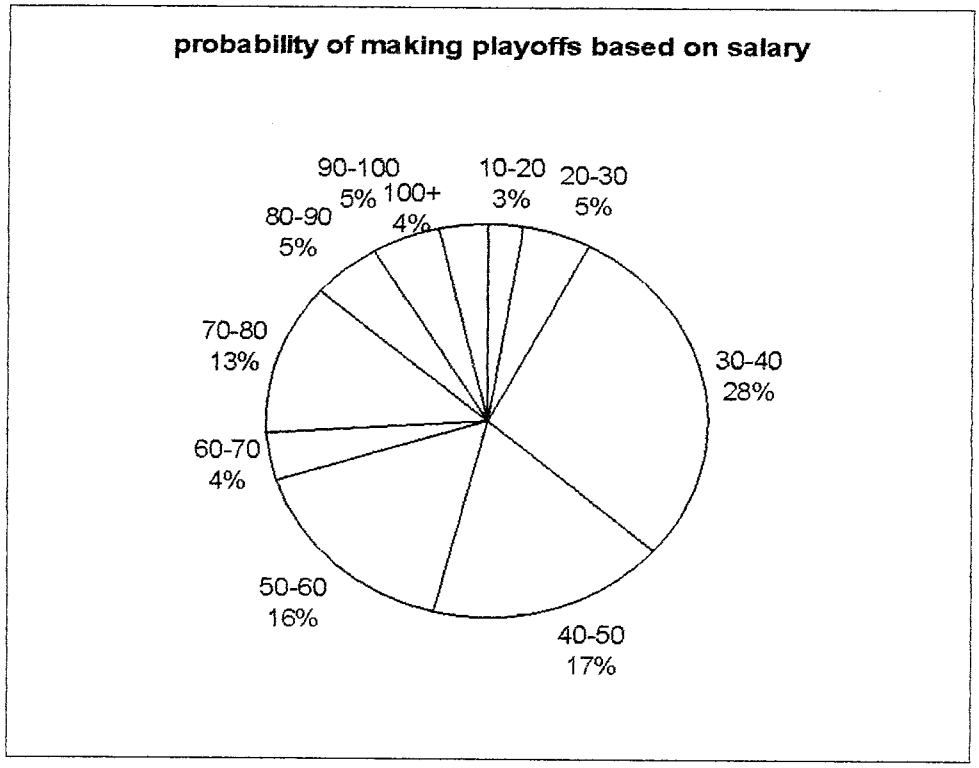
Salary Rank	# of World Series won	Probability of winning World Series based on Salary Rank
1-3	7	70%
4-6	1	10%
7-9	1	10%
10-12	0	0%
13-15	1	10%
16-18	0	0%
19-21	0	0%
21-23	0	0%
24-26	0	0%
27-29	0	0%
30+	0	0%

The following pie graph assists in solving the problem from the previous graph. Instead of using actual amounts I used the team's salary ranking within the league. This allows us to see where each team stands when spending money on players and whether is results in a higher probability of winning the World Series. From this graph you can see that 7 of the previous 10 champions were in the top 3 for total team salary. This results in a 70% chance of winning a World Series if a team places within the top 3 for team salary. Also it is interesting to note that 1 team was able to win the championship while being rank 15th in team salary.

Probability of winning World Series based on salary ranking

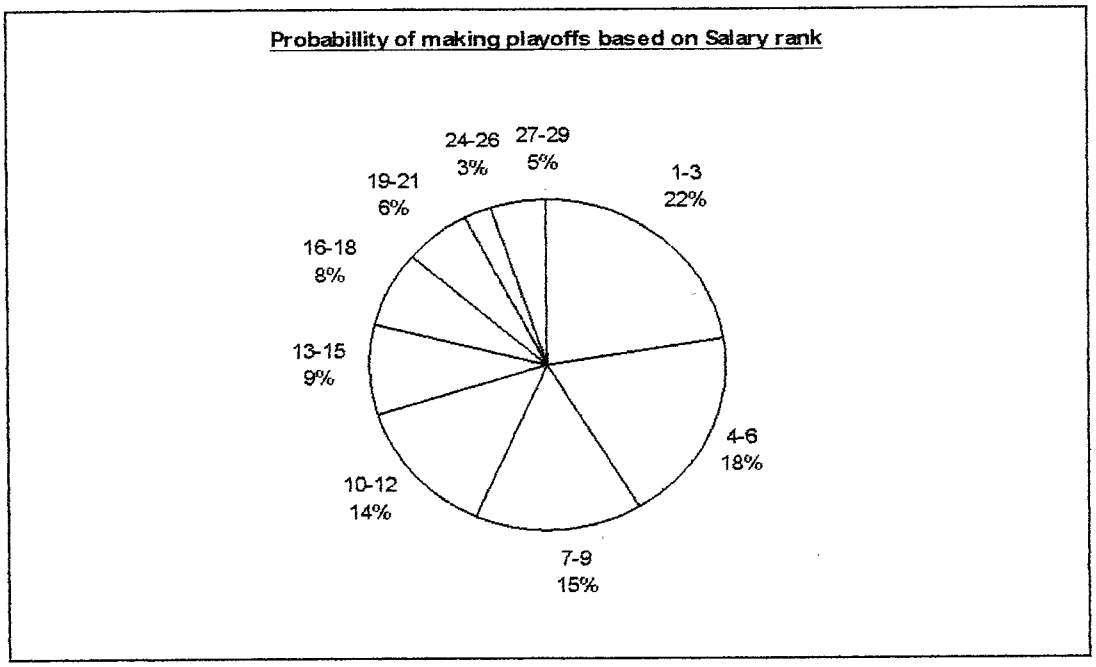


Team Salary (in millions)	# of playoff appearances	probability of making playoffs based on Salary
10-20	2	3%
20-30	4	5%
30-40	23	29%
40-50	14	18%
50-60	13	16%
60-70	3	4%
70-80	10	13%
80-90	4	5%
90-100	4	5%
100+	3	4%



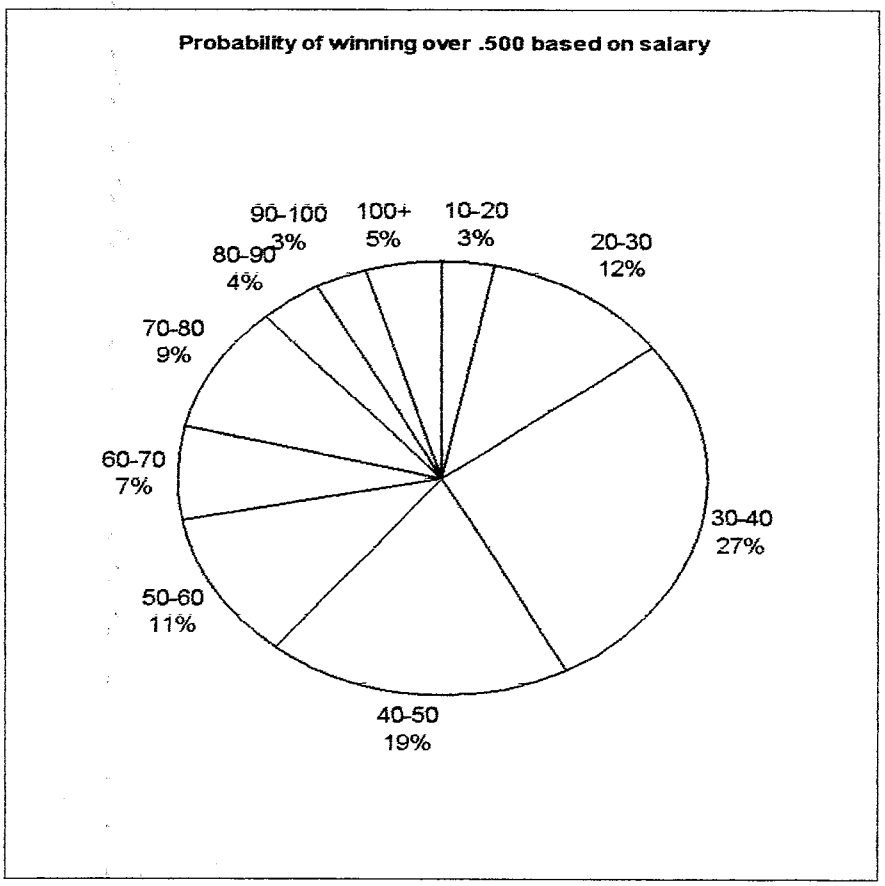
In order to get to the championships, a team must get into the playoffs. The following shows the probability of getting into the playoffs based on the amount the team had spent. I recorded which teams went to the playoffs and found the team salary for that particular year. For example, the Anaheim Angels went to the playoffs in 2002 and spent 61721667, so I placed there playoff appearance into the 60-70 million category. I did this for every year for the last 10 years except 1994 because no playoff was held that year. The graph above shows the results. This graph also suffers from the same problem in which the low team salary costs during the early 90s make it seem as though a team can spend less and still get into the playoffs.

Salary Rank	# of playoff appearances	Probabilily of making playoff based on Salary ranking
1-3	18	23%
4-6	15	19%
7-9	12	15%
10-12	11	14%
13-15	7	9%
16-18	6	8%
19-21	5	6%
21-23	0	0%
24-26	2	3%
27-29	4	5%
30+	0	0%



This graph was placed here to correct the issue of team salaries displaying false probabilities since the last graph made it seem as though spending 30-40 million would get a team into the playoffs. This graph simply shows the probability of making the playoffs based on the ranking, it allows for a much better perspective since it involves looking at what the other teams have spent as well. The graph is as expected, the higher the salary rank of a team among the league the better the chance of making the playoffs. By ranking in the top 3 within the league, a team has a 22% chance of making the playoffs. This graph helps supports my initial theory that team salary does have a large impact on making the playoffs, which was stated in the double variable analysis.

Team Salary (in millions)	# of teams with over .500 record	Probability of earning .500 record based on salary
10-20	4	3%
20-30	15	12%
30-40	35	27%
40-50	24	19%
50-60	14	11%
60-70	9	7%
70-80	12	9%
80-90	5	4%
90-100	4	3%
100+	6	5%



The above graph is displays the probability of having more wins than losses (<.500). This is important because in order to have a "winning" season, a team must win greater than half its games. This graph is identical to the others; I followed the same method of creating the probability distribution table. Again, like the others, it has the same flaw in which a low team salaries in the early 90s are making it seems as though teams can win by spending small amounts of money. This and the others graphs will be accompanied by Team Salary Ranking graphs in the final project. This will help show where each team actually stands among team salaries.

Probability of winning World Series based on Salary

Trial 1	71	Trial 36	73	Trial 71	15
Trial 2	57	Trial 37	38	Trial 72	99
Trial 3	10	Trial 38	78	Trial 73	83
Trial 4	31	Trial 39	40	Trial 74	60
Trial 5	12	Trial 40	61	Trial 75	15
Trial 6	61	Trial 41	16	Trial 76	4
Trial 7	31	Trial 42	89	Trial 77	33
Trial 8	70	Trial 43	33	Trial 78	20
Trial 9	78	Trial 44	23	Trial 79	63
Trial 10	84	Trial 45	38	Trial 80	88
Trial 11	66	Trial 46	60	Trial 81	86
Trial 12	13	Trial 47	46	Trial 82	83
Trial 13	65	Trial 48	31	Trial 83	75
Trial 14	80	Trial 49	46	Trial 84	9
Trial 15	70	Trial 50	75	Trial 85	87
Trial 16	37	Trial 51	81	Trial 86	68
Trial 17	24	Trial 52	99	Trial 87	70
Trial 18	35	Trial 53	23	Trial 88	58
Trial 19	86	Trial 54	12	Trial 89	6
Trial 20	65	Trial 55	27	Trial 90	100
Trial 21	80	Trial 56	85	Trial 91	27
Trial 22	30	Trial 57	86	Trial 92	78
Trial 23	51	Trial 58	100	Trial 93	36
Trial 24	16	Trial 59	57	Trial 94	71
Trial 25	18	Trial 60	59	Trial 95	58
Trial 26	93	Trial 61	42	Trial 96	18
Trial 27	5	Trial 62	32	Trial 97	55
Trial 28	18	Trial 63	17	Trial 98	91
Trial 29	50	Trial 64	4	Trial 99	72
Trial 30	66	Trial 65	31	Trial 100	35
Trial 31	25	Trial 66	84		
Trial 32	68	Trial 67	38		
Trial 33	62	Trial 68	76		
Trial 34	34	Trial 69	87		
Trial 35	80	Trial 70	84		

Probability of making the playoffs based on Salary

Probability of winning World series based on Salary Rank

Trial 1	71	Trial 36	61	Trial 71	54
Trial 2	19	Trial 37	72	Trial 72	53
Trial 3	10	Trial 38	50	Trial 73	56
Trial 4	15	Trial 39	10	Trial 74	11
Trial 5	52	Trial 40	20	Trial 75	61
Trial 6	35	Trial 41	7	Trial 76	3
Trial 7	85	Trial 42	15	Trial 77	9
Trial 8	5	Trial 43	15	Trial 78	53
Trial 9	99	Trial 44	91	Trial 79	46
Trial 10	16	Trial 45	25	Trial 80	91
Trial 11	55	Trial 46	40	Trial 81	51
Trial 12	60	Trial 47	28	Trial 82	61
Trial 13	89	Trial 48	70	Trial 83	22
Trial 14	8	Trial 49	94	Trial 84	85
Trial 15	50	Trial 50	36	Trial 85	92
Trial 16	66	Trial 51	10	Trial 86	33
Trial 17	90	Trial 52	16	Trial 87	43
Trial 18	97	Trial 53	9	Trial 88	12
Trial 19	38	Trial 54	5	Trial 89	21
Trial 20	42	Trial 55	43	Trial 90	80
Trial 21	33	Trial 56	26	Trial 91	54
Trial 22	71	Trial 57	70	Trial 92	34
Trial 23	39	Trial 58	37	Trial 93	8
Trial 24	52	Trial 59	37	Trial 94	52
Trial 25	74	Trial 60	19	Trial 95	71
Trial 26	38	Trial 61	56	Trial 96	50
Trial 27	24	Trial 62	70	Trial 97	95
Trial 28	97	Trial 63	77	Trial 98	54
Trial 29	21	Trial 64	47	Trial 99	66
Trial 30	31	Trial 65	56	Trial 100	24
Trial 31	26	Trial 66	35		
Trial 32	62	Trial 67	20		
Trial 33	6	Trial 68	68		
Trial 34	25	Trial 69	41		
Trial 35	47	Trial 70	85		

Probability of winning over .500 games based on salary

Trial 1	15	Trial 36	43	Trial 71	35
Trial 2	56	Trial 37	21	Trial 72	73
Trial 3	29	Trial 38	92	Trial 73	53
Trial 4	66	Trial 39	41	Trial 74	43
Trial 5	97	Trial 40	14	Trial 75	34
Trial 6	77	Trial 41	91	Trial 76	76
Trial 7	99	Trial 42	23	Trial 77	1
Trial 8	92	Trial 43	66	Trial 78	62
Trial 9	21	Trial 44	1	Trial 79	21
Trial 10	79	Trial 45	97	Trial 80	10
Trial 11	10	Trial 46	61	Trial 81	44
Trial 12	93	Trial 47	60	Trial 82	90
Trial 13	41	Trial 48	88	Trial 83	59
Trial 14	86	Trial 49	34	Trial 84	32
Trial 15	98	Trial 50	9	Trial 85	83
Trial 16	15	Trial 51	37	Trial 86	42
Trial 17	26	Trial 52	51	Trial 87	37
Trial 18	1	Trial 53	74	Trial 88	62
Trial 19	95	Trial 54	90	Trial 89	52
Trial 20	29	Trial 55	97	Trial 90	44
Trial 21	60	Trial 56	63	Trial 91	52
Trial 22	65	Trial 57	89	Trial 92	21
Trial 23	23	Trial 58	32	Trial 93	41
Trial 24	74	Trial 59	32	Trial 94	21
Trial 25	34	Trial 60	36	Trial 95	65
Trial 26	21	Trial 61	45	Trial 96	70
Trial 27	80	Trial 62	63	Trial 97	31
Trial 28	31	Trial 63	45	Trial 98	2
Trial 29	87	Trial 64	55	Trial 99	27
Trial 30	16	Trial 65	78	Trial 100	19
Trial 31	64	Trial 66	82		
Trial 32	13	Trial 67	54		
Trial 33	91	Trial 68	67		
Trial 34	49	Trial 69	62		
Trial 35	20	Trial 70	54		

Trial 1	62	Trial 36	8	Trial 71	35
Trial 2	10	Trial 37	16	Trial 72	22
Trial 3	50	Trial 38	74	Trial 73	97
Trial 4	56	Trial 39	47	Trial 74	50
Trial 5	89	Trial 40	18	Trial 75	36
Trial 6	10	Trial 41	61	Trial 76	6
Trial 7	62	Trial 42	26	Trial 77	88
Trial 8	92	Trial 43	92	Trial 78	68
Trial 9	52	Trial 44	58	Trial 79	6
Trial 10	49	Trial 45	33	Trial 80	13
Trial 11	98	Trial 46	68	Trial 81	33
Trial 12	97	Trial 47	9	Trial 82	50
Trial 13	97	Trial 48	88	Trial 83	68
Trial 14	97	Trial 49	83	Trial 84	4
Trial 15	95	Trial 50	94	Trial 85	49
Trial 16	43	Trial 51	100	Trial 86	66
Trial 17	12	Trial 52	68	Trial 87	62
Trial 18	67	Trial 53	52	Trial 88	74
Trial 19	79	Trial 54	57	Trial 89	76
Trial 20	8	Trial 55	43	Trial 90	61
Trial 21	77	Trial 56	99	Trial 91	25
Trial 22	99	Trial 57	51	Trial 92	65
Trial 23	43	Trial 58	52	Trial 93	30
Trial 24	21	Trial 59	67	Trial 94	13
Trial 25	37	Trial 60	57	Trial 95	53
Trial 26	99	Trial 61	22	Trial 96	6
Trial 27	36	Trial 62	19	Trial 97	20
Trial 28	85	Trial 63	72	Trial 98	83
Trial 29	15	Trial 64	33	Trial 99	80
Trial 30	63	Trial 65	5	Trial 100	79
Trial 31	76	Trial 66	40		
Trial 32	80	Trial 67	80		
Trial 33	25	Trial 68	76		
Trial 34	92	Trial 69	82		
Trial 35	65	Trial 70	70		

Simulation Results and Expected Value

Probability of winning World Series based on Salary

Simulation	Results	%
40-50 million, #1-37	36	36%
50-60 million, #38-50	9	9%
60-70 million, #51-75	29	29%
80-90 milllion,#76-100	26	26%

Probability of winning World series based on Salary Rank

Simulation	Results	
rank 1-3, #1-70	82	82%
rank 4-6, #71-80	7	7%
rank 7-9, #81-90	5	5%
rank 13-15, #91-100	6	6%

Probability of making the playoffs based on Salary

Simulation	Results	
10-20 million, #1-3	4	4%
20-30 million, #4-8	0	0%
30-40 million, #9-37	34	34%
40-50 million, #38-55	18	18%
50-60 million, #56-71	17	17%
60-70 million, #72-75	3	3%
70-80 milllion, #76-88	10	10%
80-90 million,#89-93	8	8%
90-100 million, #94-98	6	6%
100+ million, #98-100	0	0%

Probability of winning over .500 games based on salary

Simulation	Results	
10-20 million, #1-3	0	0%
20-30 million, #4-15	12	12%
30-40 million, #16-42	19	19%
40-50 million, #43-61	22	22%
50-60 million, #62-72	14	14%
60-70 million, #73-79	9	9%
70-80 milllion, #80-88	9	9%
80-90 million,#89-92	4	4%
90-100 million, #93-95	2	2%
100+ million, #96-100	9	9%

1) $E(X) = 85*.25+45*.37+65*.25+55*.13$
$= 21.25+16.65+16.25+7.15$
$= 61.3$

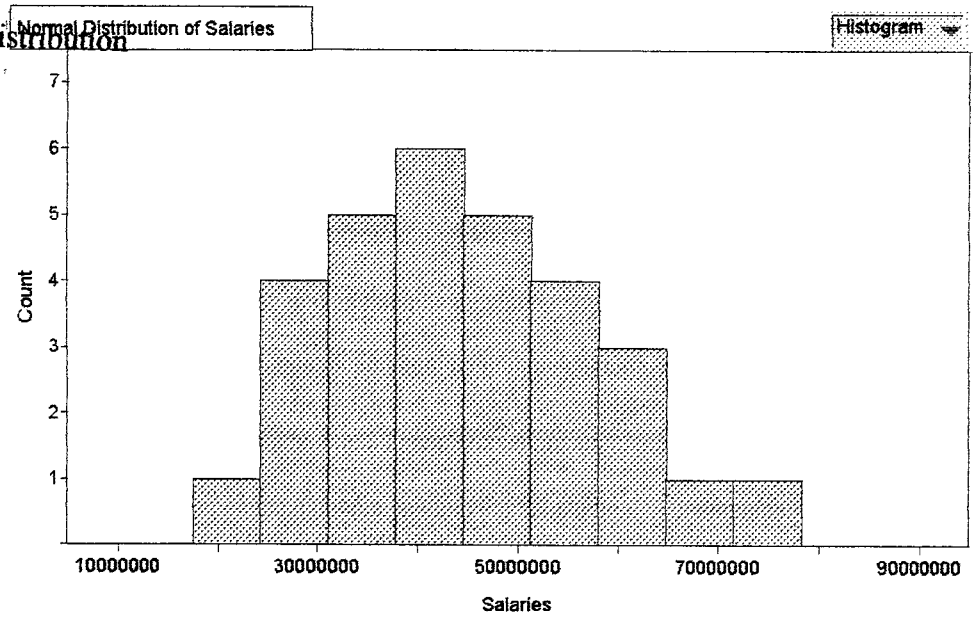
2) $E(X) = 2*.7+5*.1+8*.1+14*.1$
$= 1.4+0.5+0.8+1.4$
$= 4.1$

3) $E(X) = 15*.03+25*.05+35*.29+45*.18+55*.16+65*.04+75*.13$
$+85*.05+95*.05+100*.04$
$= .45+1.25+10.15+8.1+8.8+2.6+9.75+4.25+4.75+4$
$= 54.1$

4) $E(X) = 15*.03+25*.12+35*.27+45*.19+55*.11+65*.07+75*.09$
$+85*.04+95*.03+100*.05$
$= .45+ 3+9.45+8.55+6.05+4.55+6.75+3.4+2.85+5$
$= 50.05$

Each expected value calculation coordinates with the chart right beside it to the left, but because this is expected value I did not use the results from the simulation I used the probability distribution tables I made earlier. Calculation one means that a team is expected to spend 61.3 million dollars in order to win the World Series. Calculation two means that a team is expected to be ranked 4.1 in team salary to win the World Series. Calculation three means a team is expected to spend 54.1 million dollars to make it into the playoffs. Calculation four means a team is expected to spend 50.05 million dollars in order to have a winning season or have a record greater than .500.

Normal Distribution



Mean= 43,857,300 Standard Deviation= 12,704,200

The graph above is the normal distribution for the median salaries of each team over the last ten years. The graph shows that most of the teams have spent roughly the same amount which is appropriate for a normal distribution. The mean is apporpriate and supports the data found on the graph, 43 857 300 is very close to the middle of the graph above. Inorder to be fit a normal distribution, 68% of the data must lie within 1 standard deviation. The following calculations will prove that the data above is in fact a normal distribution.

mean-standard deviation
43 857 300- 12 704 200=31 153 100

mean+stadard deviation
43 857 300+12 704 200=56 561 500

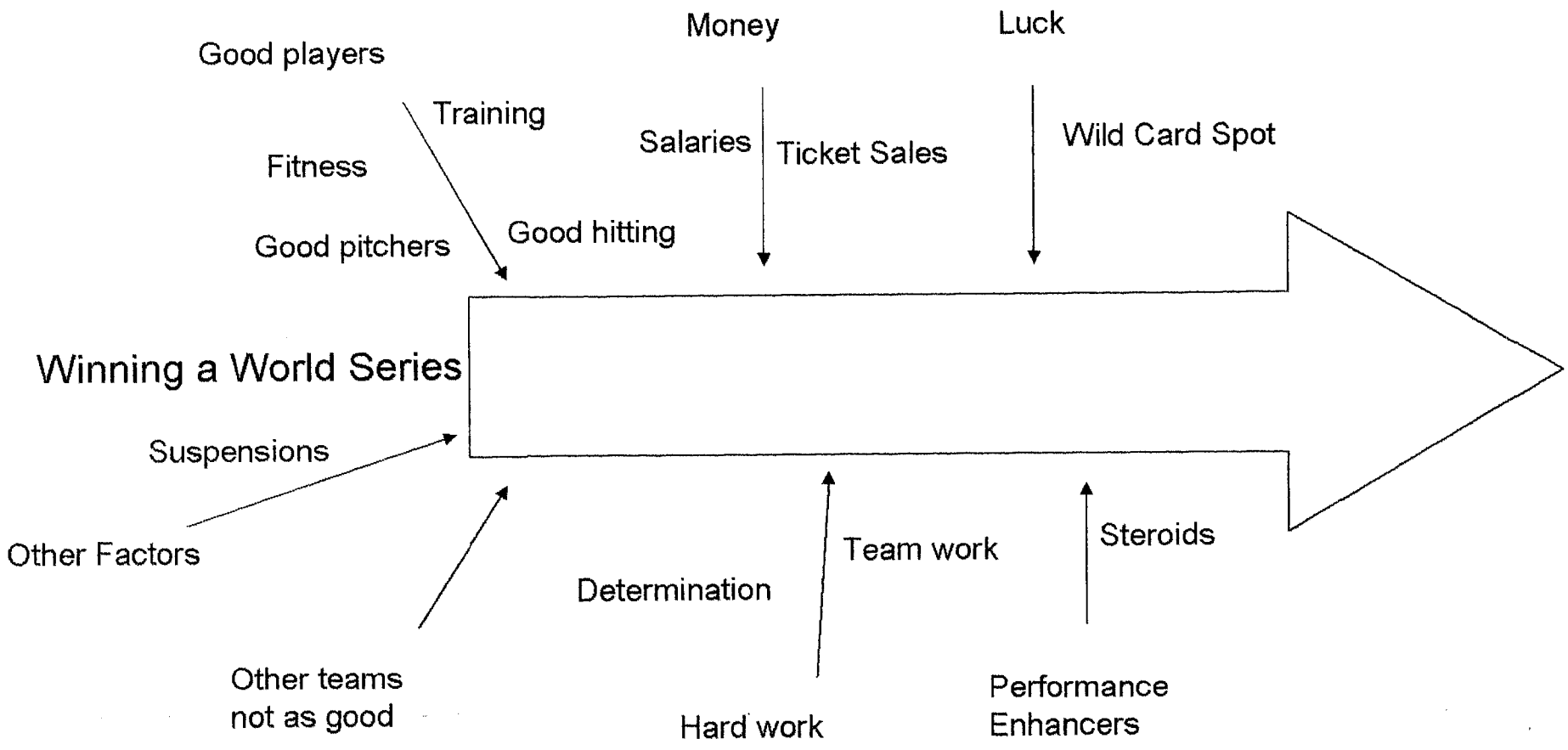
Data between 31 153 100 - 56 561 500 must represent 68%

z= (x- mean)/ stan dev
z= (31 153 100- 43857 300)/12 704 200
200
z=-1
=0.1587

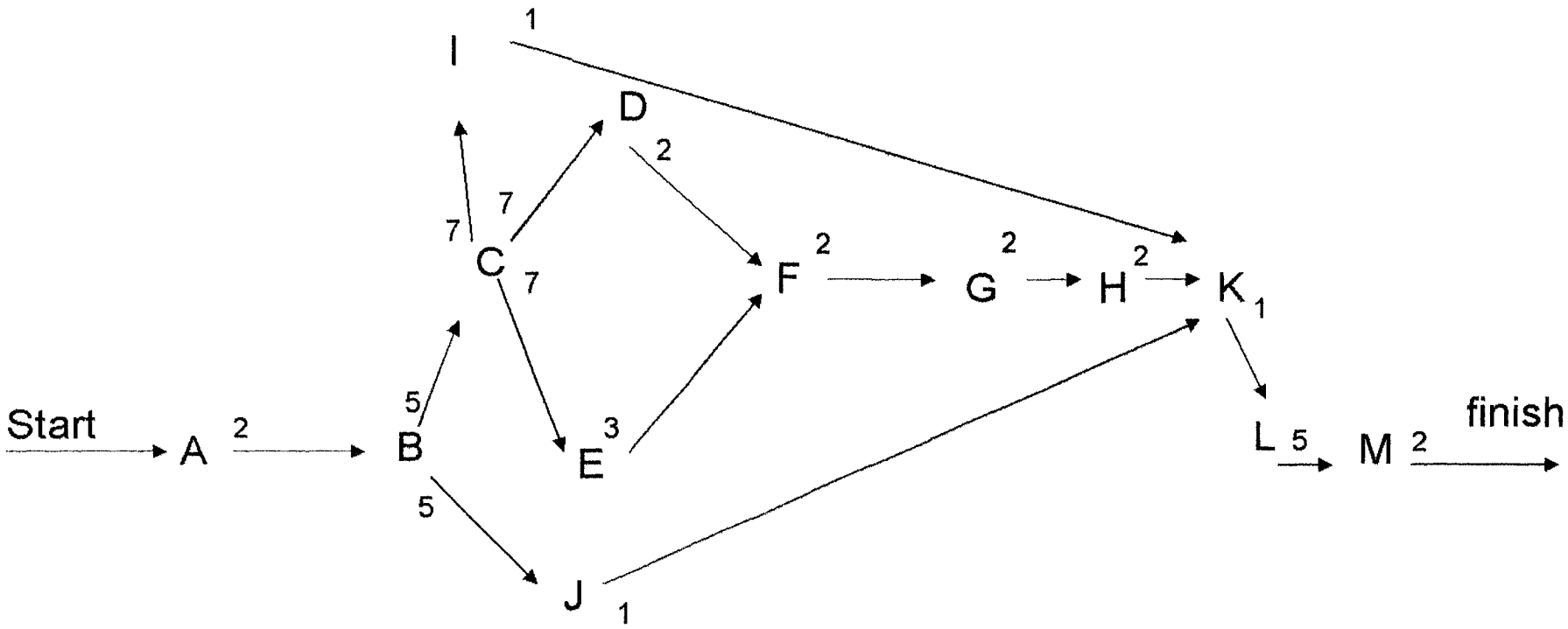
z= (x- mean)/ stan dev
z=(56 561 500 -43 857 300)/ 12 704
z= 1
=.8413

84.13-15.87 =68.26

68% of the data lies between 1 standard devation, the graph represents a normal distribution
Most of the data (68%) lies between the 15.87 percentile and 84.13 percentile.


Money
Luck
Good players
Training
Salaries
Ticket Sales
Wild Card Spot
Fitness
Good pitchers
Good hitting
Winning a World Series
Suspensions
Other Factors
Team work
Steroids
Determination
Other teams
not as good
Hard work
Performance
Enhancers

Task	Description	Duration (days)	Prerequisites
A	Mind Map	2	none
B	Specify a thesis	5	A
C	Gather Data	7	A,B
D	Single Variable Analysis	2	C
E	Double Variable Analysis	3	C
F	Written Report of Single and Double variable analysis	2	D,E
G	Probability Distribution	2	F
H	Simulation	2	G
I	Normal Distribution	1	C
J	Cause and Effect Diagram	1	B
K	Critical Path Analysis	1	I, J,H
L	Final Report (conclusions)	5	K
M	Presentation preparation	2	L



Critical Path= A, B,C, E, F,G,H, K, L, M = 31 days

Conclusion

Athletes, talented individuals that train hours and hours each day to win accomplish one goal, a championship. They provide hours and hours of endless entertainment, as a result they get paid a king's ransom. For many, the money that athletes earn would last a life time, yet owners of teams for some reason keep splurging more and more money all in hope to win the championship. In baseball, this is the World Series. Year after year, owners endlessly search for younger, faster, better players, all in hope of winning this one trophy. With cash bleeding out of their pockets, owners continue to spend away, but at what point does an owner stop? When does he/she feel they have spent enough money to have a reasonable shot at the championship? The entire purpose of my project was to find that one point.

After thorough examination of my data, I've come to several conclusions. The first being that except for the odd case, the teams that spend the most money tend to win more in regular season play. As well, these teams tend to make the playoff far more often than teams that spend less amounts of money. The question I was to answer was "Does having a higher team salary result in more wins and ultimately a World Series?" The response to the first question was already given, yes spending more results in a more wins, but the answer to the second part is no. Aside from the Yankees, the World Series champions have been almost random when factoring in the team salaries. Past World Series champions have included teams that have spent half as much as other teams, teams that spent the same amount as the rest of the league and teams that have been ranked by salary ranging from 1^{st} to 14^{th}. However, all these exceptions I explained, do not deny the fact that spending more will get you a better shot at winning the championship. Whether it helps your team get into the playoffs or lets your team advance further into the playoffs, spending more does have its advantages.

In conclusion, spending more money on players gets a team pretty far, but it simply doesn't take you all the way. Team work, dedication, strategies, training and many other factors make a team a champion. After all, before there was money, players were playing the game because they loved it.

Errors

The following are possible errors that may have occurred in my project:

- 10 years not long enough
- Some teams were non-existent, may have affected some graphs
- Canadian teams had to pay exchange rate, means they actual had to spend more
- Injuries to key players or players with big salaries could have affected the teams success resulting in a poor season but still a high team salary
- 1994 world series and playoffs were cancelled due to strike
- data from sources may be incorrect

Bibliography

http://www.sportsline.com/mlb/standings/archive/2001

http://sportsillustrated.cnn.com/baseball/mlb/standings/1999/

http://sports.espn.go.com/mlb/standings?type=reg&br=3&year=2002&column=gamesBehind&order=false&st=2

http://mlb.mlb.com/NASApp/mlb/mlb/history/postseason/mlb_ws.jsp?feature=club_champs

http://baseball1.com/c-stats.html

http://asp.usatoday.com/sports/baseball/salaries/default.aspx