Program Summary - finalPro.sas

Execution Environment

Author:	rpang4@wisc.edu
File:	/home/u39737451/ruipang/finalPro.sas
SAS Platform:	Linux LIN X64 3.10.0-693.21.1.el7.x86_64
SAS Host:	ODAWS01-USW2.ODA.SAS.COM
SAS Version:	9.04.01M6P11072018
SAS Locale:	en_US
Submission Time:	10/22/2019, 9:12:37 PM
Browser Host:	HOST-9-142.WIMASTS.MADISON.WI.US.CLIENTS.PAVLOVMEDIA.NET
User Agent:	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/68.0.3440.106 Safari/537.36
Application Server:	ODAMID01-PROD-US.ODA.SAS.COM

Code: finalPro.sas

```
/*
 Rui Pang 07/11/20190
 *I am a big fan of basketball. In addition, I noticed that powerful team have to
 * have a powerful weapon to win the game, like rebounds, 3-points and so on.
 * I concern about higher 3-Point attempting rate whether are the factor for winning game or not.
 * That analyzation could explains the current tendency for the basketball.
 * Here are variables I will use for analyzation:
 * 3PAr(Percentage of FG Attempts from 3-Point Range)
 * W(Total game wins)
 */


/*
This data set recorded average team statistics and informations
for whole 30 observations(NBA teams) in season 2018-2019 where I found in basketball-reference.com.
This data set contains 25 variables to display the overall status for each NBA teams.
Each team will have 82 games for whole season.
*/
proc datasets library=work kill noprint;
    run;
quit;

filename mynba url
    "https://raw.githubusercontent.com/ruipang123/stat479/master/nnba.csv"
    termstr=lf;

/*
This is a part that I import team shooting data from 2018-2019 season .
*/
filename myshoot url
    "https://raw.githubusercontent.com/ruipang123/stat479/master/Shoot" termstr=lf;

/*
This is a part that importing NBA League Averages Data for each season.
*/
filename myseason url
    "https://raw.githubusercontent.com/ruipang123/stat479/master/season%20summary"
    termstr=lf;

/*
 * This data step helps me to create a data set called nbafinal.
 * Nbafinal contains 28 variables and 30 obervations.
 * In addition, I created a new variable caller winrate which using the number of
 * winning games divided by total number of games(82) to have the rate of winning
 * for each team.
 */
data nbafinal;
    infile MYNBA delimiter=',' MISSOVER DSD lrecl=32767 firstobs=2;
    informat Rk best32. Team $23. Age best32.
    W best32.  L best32. PW best32.
```

```
    PL best32. MOV best32. SOS best32.
    SRS best32. ORtg best32. DRtg best32.
    NRtg best32. Pace best32. FTr best32.
    threeAr best32. TS best32. OeFG best32.
    OTOV best32. ORB best32. OFT best32.
    DeFG best32. DTOV best32. DRB best32.
    DFT best32. Arena $26. Attend best32. AttendG best32.;
    input Rk team $ age W L PW PL MOV SOS SRS ORtg DRtg NRtg Pace FTr threeAr TS
        OeFG OTOV ORB OFT DeFG DTOV DRB DFT Arena $ Attend AttendG;
    Winrate=w/82;
    label w="Wins" l="Lost" team="Team Name";
run;

/*
 * This data step creates a date set called shoot for each team's
 * shooting statistic from 2018-2019 season.
 * Shoot contains 25 variables and 30 observations(NBA teams).
 */
data WORK.SHOOT;
    infile MYSHOOT delimiter=',' MISSOVER DSD lrecl=32767 firstobs=2;
    input Rk :2. Team :$23. G :5. MP :5.2 FG :5.2 FGA :5.2 FGper :5. threeP :5.2
        threePA :5.2 threePP :5.2 twoP :5.2 twoPA :5.2 PPN :5.2 FT :5.2 FTA :5.2
        FTP :5.2 ORB :5.2 DRB :5.2 TRB :5.2 AST :5.2 STL :5.2 BLK :5.2 TOV :5.2
        PF :5.2 PTS :5.2;
run;


/*
 * This data step creates a date set called season for league statistic from each season(2010-2019).
 * Season contains 4 variables and 9 observations(NBA teams).
 */

data WORK.SEASON;
    infile MYSEASON delimiter=',' MISSOVER DSD lrecl=32767 firstobs=2;
    input Rk :2. Season  :$7. Lg  :$3. Age :4.1 Ht :$3. Wt :3. Game :5. MP :5.1
        FG :4.1 FGA :4.1 ThrPFG :4.1 Attempts :4.1 FT :4.1 FTA :4.1 ORB :4.1 DRB :4.1
        TRB :4.1 AST :4.1 STL :3.1 BLK :3.1 TOV :4.1 PF :4.1 PTS :5.1 FGper :5.3
        threepp :5.3 FTper :5.3 Pace :5.1 eFG :5.3 TOVPER :4.1 ORBper :4.1 FTPA 5.3
        ORtg :5.1;
    keep Season ThrPFG Attempts PTS;
run;

/*
 * This proc sort step will help me to
 * sort season as increasing order.
 * That could help me to make the plot more
 * clearly.
 */
proc sort data=work.season out=sortsea;
    by season;
run;

/*
 * This proc step creates a graph that contains vertical bars and series lines
 * for league statistic from each season(2010-2019).
 * Seasons are my independent variable which is x-axis.
 * Vertical Bar's dependent variable is average game points.
 * Blue Series line's  dependent variable is average number of 3-point filed goal attempting.
 * Red Series line's  dependent variable is average number of 3-point filed goals.
 * Left side of y-axis is points per game which is the measurement for vertical bar.
 * The other side is number of three-point shot that is the measurement for series lines.
 * Red line with circle marks represent the number of three-point field goals per game.
 * Blue line with star marks represent the number of three-point shot attempting.
 */
proc sgplot data=sortsea;
    title 'League Season summary';
    footnote 'Summary data form basketball-reference.com';
    vbarparm category=season response=PTS/ datalabel datalabelpos=data;
    series x=season y=Attempts/ y2axis markers datalabel=Attempts
        lineattrs=(color=blue) markerattrs=(symbol=Star color=blue) datalabelpos=top;
    series x=season y=thrPFG/ y2axis markers datalabel=thrPFG
```

```
        lineattrs=(color=red) markerattrs=(color=red) datalabelpos=top;
    yaxis values=(85 to 115 by 1) label='Points Per Game';
    y2axis values=(0 to 34 by 1) label='Number of 3-Point Shot';
    inset 'Attempts of 3-Point per game: blue star'
        /title='Reference line represent:' titleattrs=(weight=bold) valuealign=center
        position=topleft textattrs=(color=blue);
    inset '   ' '    ' '3-Point Field Goals Per Game: red circle'/
        position=topleft textattrs=(color=red);
run;

title;
footnote;

/*
 * These two proc sort steps are the preparation for my next data merge step.
 * Sort nbafinal and shoot by team name.
 */
proc sort data=nbafinal out=realnbas;
    by team;
run;

proc sort data=shoot out=sortshoot;
    by team;
run;

/*
 * This data step merged realnbas and sortshoot together,
 * and creates a new data set called bignba.
 * Bigba has a new variables called playoff.
 * I used if statement to separated team to two groups.
 * I created threpp to represent the percentage of 3-point field goals
 * by multiply 100 from the decimal representation.
 */
data Bignba;
    merge realnbas sortshoot;
    by team;
    threpp=threepp*100;
    if find (team, '*') then
        playoff='Yes';
    else
        playoff='No';
    keep team age W L Winrate playoff threpp threePA threepp threep;
run;

/*
 * These proc sort steps is the preparation for my next plot step.
 * Sort bignba by winning rate, because I will use the winning rate as
 * independent variables to make scatter plot.
 */

proc sort data=bignba out=sortwin;
    by descending winrate;
run;

/*
 * This proc mean step gives me the data that
 * I will need for scatter graph's reference line.
 * This proc step calculated the mean percentage
 * of 3-point field goals per game by group of playoff.
 */
proc means data=bignba;
    var threpp;
    class playoff;
    output out=meanthre;
run;

/*
 * These macro variables store the
 * mean percentage of 3-point field
 * goals per game by group of playoff
 * and non-playoff teams.
```

```
 */
%let playoff = 36.30625;
%let nplayoff = 34.68571;

/*
 * This proc step creates a graph that contains scatter points
 * for team statistic from each last season(2018-2019).
 * Winning rate is my independent variable which is x-axis.
 * Average percentage of 3-point field goals per game is my dependent variable which is y-axis.
 * Green line represents the mean percentage of 3-point field goals per game for playoff teams.
 * Black line represents the mean percentage of 3-point field goals per game for non-playoff teams.
 */
proc sgplot data=sortwin;
    title 'Team Season summary (2018-2019)';
    footnote 'Summary data form basketball-reference.com';
    scatter x=winrate y=threpp/ group=playoff datalabel=team;
    yaxis label='Average 3-point field goal(%) per game';
    refline &nplayoff / axis=y lineattrs=(color=black) label=("&nplayoff");
    refline &playoff / axis=y lineattrs=(color=green) label=("&playoff");
    inset 'Average 3-point field goal(%)' /title='Reference line represent:'
        titleattrs=(weight=bold) valuealign=center position=topleft
        textattrs=(color=blue);
    inset '   ' '    ' ' Non-playoff team: black'/ position=topleft
        textattrs=(color=black);
    inset '   ' '    ' ' ' ' Playoff team: green'/ position=topleft
        textattrs=(color=green);
run;

/*
 * The last date step gives a data set follows a
 * decreasing order on winning rate. In addition,
 * I created a Rank variable to see the ranking of
 * the winning rate.
 */
data ss;
    set sortwin;
    Rank=_n_;
run;

/*
 * This proc print step out put the data set that
 * contains all the variables that I used for this
 * project. I change them to the correct format and
 * made some label for these variables.
 */
proc print data=ss noobs split='*';
    id rank;
    var team age W L Winrate playoff threePA threepp threep;
    label w="Win" l='Lost' threepp='Percentage 3-point*field goals'
        threePA='3-point*Attempts' threep="Numbers of 3-point*field goals";
    format threepp PERCENT5.1;
run;

title;
footnote;
```

Log: finalPro.sas

Notes (73)

```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
70
71         /*
72          Rui Pang 07/11/20190
73          *I am a big fan of basketball. In addition, I noticed that powerful team have to
74          * have a powerful weapon to win the game, like rebounds, 3-points and so on.
75          * I concern about higher 3-Point attempting rate whether are the factor for winning game or not.
76          * That analyzation could explains the current tendency for the basketball.
77          * Here are variables I will use for analyzation:
78          * 3PAr(Percentage of FG Attempts from 3-Point Range)
```

```
79          * W(Total game wins)
80          */
81
82
83         /*
84         This data set recorded average team statistics and informations
85         for whole 30 observations(NBA teams) in season 2018-2019 where I found in basketball-reference.com.
86         This data set contains 25 variables to display the overall status for each NBA teams.
87         Each team will have 82 games for whole season.
88         */
89         proc datasets library=work kill noprint;
NOTE: Deleting WORK.REGSTRY (memtype=ITEMSTOR).
NOTE: File WORK.REGSTRY (memtype=ITEMSTOR) cannot be deleted because it is in use.
NOTE: Deleting WORK.SASGOPT (memtype=CATALOG).
NOTE: File WORK.SASGOPT (memtype=CATALOG) cannot be deleted because it is in use.
NOTE: Deleting WORK.SASMAC1 (memtype=CATALOG).
NOTE: File WORK.SASMAC1 (memtype=CATALOG) cannot be deleted because it is in use.
NOTE: Deleting WORK.SASMAC2 (memtype=CATALOG).
NOTE: Deleting WORK.SASMAC3 (memtype=CATALOG).
NOTE: Deleting WORK.SASMAC4 (memtype=CATALOG).
NOTE: Deleting WORK.SASMAC5 (memtype=CATALOG).
NOTE: Deleting WORK.SASMAC6 (memtype=CATALOG).
NOTE: Deleting WORK.SASMAC7 (memtype=CATALOG).
NOTE: Deleting WORK.SASMAC8 (memtype=CATALOG).
NOTE: Deleting WORK.SASMAC9 (memtype=CATALOG).
NOTE: Deleting WORK.SASMACR (memtype=CATALOG).
NOTE: File WORK.SASMACR (memtype=CATALOG) cannot be deleted because it is in use.
90         run;

91         quit;

NOTE: PROCEDURE DATASETS used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              553.68k
      OS Memory           27560.00k
      Timestamp           10/23/2019 02:12:33 AM
      Step Count                        24  Switch Count  16
      Page Faults                       0
      Page Reclaims                     210
      Page Swaps                        0
      Voluntary Context Switches        89
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           8


92
93         filename mynba url
94         "https://raw.githubusercontent.com/ruipang123/stat479/master/nnba.csv"
95         termstr=lf;
96
97         /*
98         This is a part that I import team shooting data from 2018-2019 season .
99         */
100        filename myshoot url
101        "https://raw.githubusercontent.com/ruipang123/stat479/master/Shoot" termstr=lf;
102
103        /*
104        This is a part that importing NBA League Averages Data for each season.
105        */
106        filename myseason url
107        "https://raw.githubusercontent.com/ruipang123/stat479/master/season%20summary"
108        termstr=lf;
109
110        /*
111         * This data step helps me to create a data set called nbafinal.
112         * Nbafinal contains 28 variables and 30 obervations.
113         * In addition, I created a new variable caller winrate which using the number of
114         * winning games divided by total number of games(82) to have the rate of winning
115         * for each team.
116         */
117        data nbafinal;
118        infile MYNBA delimiter=',' MISSOVER DSD lrecl=32767 firstobs=2;
119        informat Rk best32. Team $23. Age best32.
120        W best32.  L best32. PW best32.
121        PL best32. MOV best32. SOS best32.
122        SRS best32. ORtg best32. DRtg best32.
123        NRtg best32. Pace best32. FTr best32.
```

```
124         threeAr best32. TS best32. OeFG best32.
125         OTOV best32. ORB best32. OFT best32.
126         DeFG best32. DTOV best32. DRB best32.
127         DFT best32. Arena $26. Attend best32. AttendG best32.;
128         input Rk team $ age W L PW PL MOV SOS SRS ORtg DRtg NRtg Pace FTr threeAr TS
129         OeFG OTOV ORB OFT DeFG DTOV DRB DFT Arena $ Attend AttendG;
130         Winrate=w/82;
131         label w="Wins" l="Lost" team="Team Name";
132         run;

NOTE: The infile MYNBA is:
      Filename=https://raw.githubusercontent.com/ruipang123/stat479/master/nnba.csv,
      Local Host Name=odaws04-prod-us,
      Local Host IP addr=10.249.126.101,
      Service Hostname Name=raw.githubusercontent.com,
      Service IP addr=151.101.52.133,
      Service Name=N/A,Service Portno=443,
      Lrecl=32767,Recfm=Variable

NOTE: 30 records were read from the infile MYNBA.
      The minimum record length was 150.
      The maximum record length was 171.
NOTE: The data set WORK.NBAFINAL has 30 observations and 29 variables.
NOTE: DATA statement used (Total process time):
      real time           0.22 seconds
      user cpu time       0.03 seconds
      system cpu time     0.01 seconds
      memory              2758.96k
      OS Memory           29864.00k
      Timestamp           10/23/2019 02:12:34 AM
      Step Count                        25  Switch Count  6
      Page Faults                       0
      Page Reclaims                     1235
      Page Swaps                        0
      Voluntary Context Switches        44
      Involuntary Context Switches      13
      Block Input Operations            0
      Block Output Operations           264


133
134         /*
135          * This data step creates a date set called shoot for each team's
136          * shooting statistic from 2018-2019 season.
137          * Shoot contains 25 variables and 30 observations(NBA teams).
138          */
139         data WORK.SHOOT;
140         infile MYSHOOT delimiter=',' MISSOVER DSD lrecl=32767 firstobs=2;
141         input Rk :2. Team :$23. G :5. MP :5.2 FG :5.2 FGA :5.2 FGper :5. threeP :5.2
142         threePA :5.2 threePP :5.2 twoP :5.2 twoPA :5.2 PPN :5.2 FT :5.2 FTA :5.2
143         FTP :5.2 ORB :5.2 DRB :5.2 TRB :5.2 AST :5.2 STL :5.2 BLK :5.2 TOV :5.2
144         PF :5.2 PTS :5.2;
145         run;

NOTE: The infile MYSHOOT is:
      Filename=https://raw.githubusercontent.com/ruipang123/stat479/master/Shoot,
      Local Host Name=odaws04-prod-us,
      Local Host IP addr=10.249.126.101,
      Service Hostname Name=raw.githubusercontent.com,
      Service IP addr=151.101.52.133,
      Service Name=N/A,Service Portno=443,
      Lrecl=32767,Recfm=Variable

NOTE: 30 records were read from the infile MYSHOOT.
      The minimum record length was 126.
      The maximum record length was 138.
NOTE: The data set WORK.SHOOT has 30 observations and 25 variables.
NOTE: DATA statement used (Total process time):
      real time           0.18 seconds
      user cpu time       0.02 seconds
      system cpu time     0.00 seconds
      memory              2120.93k
      OS Memory           31656.00k
      Timestamp           10/23/2019 02:12:34 AM
      Step Count                        26  Switch Count  6
      Page Faults                       0
      Page Reclaims                     480
      Page Swaps                        0
      Voluntary Context Switches        44
      Involuntary Context Switches      0
```

```
      Block Input Operations           0
      Block Output Operations          264


146
147
148        /*
149         * This data step creates a date set called season for league statistic from each season(2010-2019).
150         * Season contains 4 variables and 9 observations(NBA teams).
151         */
152
153        data WORK.SEASON;
154        infile MYSEASON delimiter=',' MISSOVER DSD lrecl=32767 firstobs=2;
155        input Rk :2. Season  :$7. Lg  :$3. Age :4.1 Ht :$3. Wt :3. Game :5. MP :5.1
156        FG :4.1 FGA :4.1 ThrPFG :4.1 Attempts :4.1 FT :4.1 FTA :4.1 ORB :4.1 DRB :4.1
157        TRB :4.1 AST :4.1 STL :3.1 BLK :3.1 TOV :4.1 PF :4.1 PTS :5.1 FGper :5.3
158        threepp :5.3 FTper :5.3 Pace :5.1 eFG :5.3 TOVPER :4.1 ORBper :4.1 FTPA 5.3
159        ORtg :5.1;
160        keep Season ThrPFG Attempts PTS;
161        run;

NOTE: The infile MYSEASON is:
      Filename=https://raw.githubusercontent.com/ruipang123/stat479/master/season%20summary,
      Local Host Name=odaws04-prod-us,
      Local Host IP addr=10.249.126.101,
      Service Hostname Name=raw.githubusercontent.com,
      Service IP addr=151.101.52.133,
      Service Name=N/A,Service Portno=443,
      Lrecl=32767,Recfm=Variable

NOTE: Invalid data for FTPA in line 2 149-153.
RULE:     ----+----1----+----2----+----3----+----4----+----5----+----6----+----7----+----8----+----9----+----0
2         1,2018-19,NBA,26.3,6-7,218,1230,241.6,41.1,89.2,11.4,32.0,17.7,23.1,10.3,34.8,45.2,24.6,7.6,5.0,14.1
     101  ,20.9,111.2,.461,.355,.766,100.0,.524,12.4,22.9,.198,110.4 158
Rk=1 Season=2018-19 Lg=NBA Age=26.3 Ht=6-7 Wt=218 Game=1230 MP=241.6 FG=41.1 FGA=89.2 ThrPFG=11.4 Attempts=32 FT=17.7 FTA=23.1
ORB=10.3 DRB=34.8 TRB=45.2 AST=24.6 STL=7.6 BLK=5 TOV=14.1 PF=20.9 PTS=111.2 FGper=0.461 threepp=0.355 FTper=0.766 Pace=100
eFG=0.524 TOVPER=12.4 ORBper=22.9 FTPA=. ORtg=110.4 _ERROR_=1 _N_=1
NOTE: Invalid data for FTPA in line 3 147-151.
3         2,2017-18,NBA,26.4,6-7,219,1230,241.4,39.6,86.1,10.5,29.0,16.6,21.7,9.7,33.8,43.5,23.2,7.7,4.8,14.3,
     101  19.9,106.3,.460,.362,.767,97.3,.521,13.0,22.3,.193,108.6 156
Rk=2 Season=2017-18 Lg=NBA Age=26.4 Ht=6-7 Wt=219 Game=1230 MP=241.4 FG=39.6 FGA=86.1 ThrPFG=10.5 Attempts=29 FT=16.6 FTA=21.7
ORB=9.7 DRB=33.8 TRB=43.5 AST=23.2 STL=7.7 BLK=4.8 TOV=14.3 PF=19.9 PTS=106.3 FGper=0.46 threepp=0.362 FTper=0.767 Pace=97.3
eFG=0.521 TOVPER=13 ORBper=22.3 FTPA=. ORtg=108.6 _ERROR_=1 _N_=2
NOTE: Invalid data for FTPA in line 4 147-151.
4         3,2016-17,NBA,26.6,6-7,220,1230,241.6,39.0,85.4,9.7,27.0,17.8,23.1,10.1,33.4,43.5,22.6,7.7,4.7,14.0,
     101  19.9,105.6,.457,.358,.772,96.4,.514,12.7,23.3,.209,108.8 156
Rk=3 Season=2016-17 Lg=NBA Age=26.6 Ht=6-7 Wt=220 Game=1230 MP=241.6 FG=39 FGA=85.4 ThrPFG=9.7 Attempts=27 FT=17.8 FTA=23.1 ORB=10.1
DRB=33.4 TRB=43.5 AST=22.6 STL=7.7 BLK=4.7 TOV=14 PF=19.9 PTS=105.6 FGper=0.457 threepp=0.358 FTper=0.772 Pace=96.4 eFG=0.514
TOVPER=12.7 ORBper=23.3 FTPA=. ORtg=108.8 _ERROR_=1 _N_=3
NOTE: Invalid data for FTPA in line 5 147-151.
5         4,2015-16,NBA,26.7,6-7,221,1230,241.8,38.2,84.6,8.5,24.1,17.7,23.4,10.4,33.3,43.8,22.3,7.8,5.0,14.4,
     101  20.3,102.7,.452,.354,.757,95.8,.502,13.2,23.8,.209,106.4 156
Rk=4 Season=2015-16 Lg=NBA Age=26.7 Ht=6-7 Wt=221 Game=1230 MP=241.8 FG=38.2 FGA=84.6 ThrPFG=8.5 Attempts=24.1 FT=17.7 FTA=23.4
ORB=10.4 DRB=33.3 TRB=43.8 AST=22.3 STL=7.8 BLK=5 TOV=14.4 PF=20.3 PTS=102.7 FGper=0.452 threepp=0.354 FTper=0.757 Pace=95.8
eFG=0.502 TOVPER=13.2 ORBper=23.8 FTPA=. ORtg=106.4 _ERROR_=1 _N_=4
NOTE: Invalid data for FTPA in line 6 147-151.
6         5,2014-15,NBA,26.7,6-7,222,1230,242.0,37.5,83.6,7.8,22.4,17.1,22.8,10.9,32.4,43.3,22.0,7.7,4.8,14.4,
     101  20.2,100.0,.449,.350,.750,93.9,.496,13.3,25.1,.205,105.6 156
Rk=5 Season=2014-15 Lg=NBA Age=26.7 Ht=6-7 Wt=222 Game=1230 MP=242 FG=37.5 FGA=83.6 ThrPFG=7.8 Attempts=22.4 FT=17.1 FTA=22.8
ORB=10.9 DRB=32.4 TRB=43.3 AST=22 STL=7.7 BLK=4.8 TOV=14.4 PF=20.2 PTS=100 FGper=0.449 threepp=0.35 FTper=0.75 Pace=93.9 eFG=0.496
TOVPER=13.3 ORBper=25.1 FTPA=. ORtg=105.6 _ERROR_=1 _N_=5
NOTE: Invalid data for FTPA in line 7 147-151.
RULE:     ----+----1----+----2----+----3----+----4----+----5----+----6----+----7----+----8----+----9----+----0
7         6,2013-14,NBA,26.5,6-7,223,1230,242.0,37.7,83.0,7.7,21.5,17.8,23.6,10.9,31.8,42.7,22.0,7.7,4.7,14.6,
     101  20.7,101.0,.454,.360,.756,93.9,.501,13.6,25.5,.215,106.6 156
Rk=6 Season=2013-14 Lg=NBA Age=26.5 Ht=6-7 Wt=223 Game=1230 MP=242 FG=37.7 FGA=83 ThrPFG=7.7 Attempts=21.5 FT=17.8 FTA=23.6 ORB=10.9
DRB=31.8 TRB=42.7 AST=22 STL=7.7 BLK=4.7 TOV=14.6 PF=20.7 PTS=101 FGper=0.454 threepp=0.36 FTper=0.756 Pace=93.9 eFG=0.501
TOVPER=13.6 ORBper=25.5 FTPA=. ORtg=106.6 _ERROR_=1 _N_=6
NOTE: Invalid data for FTPA in line 8 146-150.
8         7,2012-13,NBA,26.7,6-7,223,1229,241.9,37.1,82.0,7.2,20.0,16.7,22.2,11.2,31.0,42.1,22.1,7.8,5.1,14.6,
     101  19.8,98.1,.453,.359,.753,92.0,.496,13.7,26.5,.204,105.8 155
Rk=7 Season=2012-13 Lg=NBA Age=26.7 Ht=6-7 Wt=223 Game=1229 MP=241.9 FG=37.1 FGA=82 ThrPFG=7.2 Attempts=20 FT=16.7 FTA=22.2 ORB=11.2
DRB=31 TRB=42.1 AST=22.1 STL=7.8 BLK=5.1 TOV=14.6 PF=19.8 PTS=98.1 FGper=0.453 threepp=0.359 FTper=0.753 Pace=92 eFG=0.496
TOVPER=13.7 ORBper=26.5 FTPA=. ORtg=105.8 _ERROR_=1 _N_=7
NOTE: Invalid data for FTPA in line 9 145-149.
9         8,2011-12,NBA,26.6,6-7,223,990,241.9,36.5,81.4,6.4,18.4,16.9,22.5,11.4,30.8,42.2,21.0,7.7,5.1,14.6,1
     101  9.6,96.3,.448,.349,.752,91.3,.487,13.8,27.0,.208,104.6 154
Rk=8 Season=2011-12 Lg=NBA Age=26.6 Ht=6-7 Wt=223 Game=990 MP=241.9 FG=36.5 FGA=81.4 ThrPFG=6.4 Attempts=18.4 FT=16.9 FTA=22.5
ORB=11.4 DRB=30.8 TRB=42.2 AST=21 STL=7.7 BLK=5.1 TOV=14.6 PF=19.6 PTS=96.3 FGper=0.448 threepp=0.349 FTper=0.752 Pace=91.3
eFG=0.487 TOVPER=13.8 ORBper=27 FTPA=. ORtg=104.6 _ERROR_=1 _N_=8
```

```
NOTE: Invalid data for FTPA in line 10 146-150.
10        9,2010-11,NBA,26.6,6-7,223,1230,241.9,37.2,81.2,6.5,18.0,18.6,24.4,10.9,30.5,41.4,21.5,7.3,4.9,14.3,
     101  20.7,99.6,.459,.358,.763,92.1,.498,13.4,26.4,.229,107.3 155
Rk=9 Season=2010-11 Lg=NBA Age=26.6 Ht=6-7 Wt=223 Game=1230 MP=241.9 FG=37.2 FGA=81.2 ThrPFG=6.5 Attempts=18 FT=18.6 FTA=24.4
ORB=10.9 DRB=30.5 TRB=41.4 AST=21.5 STL=7.3 BLK=4.9 TOV=14.3 PF=20.7 PTS=99.6 FGper=0.459 threepp=0.358 FTper=0.763 Pace=92.1
eFG=0.498 TOVPER=13.4 ORBper=26.4 FTPA=. ORtg=107.3 _ERROR_=1 _N_=9
NOTE: 9 records were read from the infile MYSEASON.
      The minimum record length was 154.
      The maximum record length was 158.
NOTE: The data set WORK.SEASON has 9 observations and 4 variables.
NOTE: DATA statement used (Total process time):
      real time           0.21 seconds
      user cpu time       0.03 seconds
      system cpu time     0.00 seconds
      memory              710.40k
      OS Memory           31656.00k
      Timestamp           10/23/2019 02:12:34 AM
      Step Count                        27  Switch Count  6
      Page Faults                       0
      Page Reclaims                     134
      Page Swaps                        0
      Voluntary Context Switches        46
      Involuntary Context Switches      4
      Block Input Operations            0
      Block Output Operations           264


162
163       /*
164        * This proc sort step will help me to
165        * sort season as increasing order.
166        * That could help me to make the plot more
167        * clearly.
168        */
169       proc sort data=work.season out=sortsea;
170       by season;
171       run;

NOTE: There were 9 observations read from the data set WORK.SEASON.
NOTE: The data set WORK.SORTSEA has 9 observations and 4 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.01 seconds
      system cpu time     0.00 seconds
      memory              1199.68k
      OS Memory           32432.00k
      Timestamp           10/23/2019 02:12:34 AM
      Step Count                        28  Switch Count  2
      Page Faults                       0
      Page Reclaims                     228
      Page Swaps                        0
      Voluntary Context Switches        12
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           272


172
173       /*
174        * This proc step creates a graph that contains vertical bars and series lines
175        * for league statistic from each season(2010-2019).
176        * Seasons are my independent variable which is x-axis.
177        * Vertical Bar's dependent variable is average game points.
178        * Blue Series line's  dependent variable is average number of 3-point filed goal attempting.
179        * Red Series line's  dependent variable is average number of 3-point filed goals.
180        * Left side of y-axis is points per game which is the measurement for vertical bar.
181        * The other side is number of three-point shot that is the measurement for series lines.
182        * Red line with circle marks represent the number of three-point field goals per game.
183        * Blue line with star marks represent the number of three-point shot attempting.
184        */
185       proc sgplot data=sortsea;
186       title 'League Season summary';
187       footnote 'Summary data form basketball-reference.com';
188       vbarparm category=season response=PTS/ datalabel datalabelpos=data;
189       series x=season y=Attempts/ y2axis markers datalabel=Attempts
190       lineattrs=(color=blue) markerattrs=(symbol=Star color=blue) datalabelpos=top;
191       series x=season y=thrPFG/ y2axis markers datalabel=thrPFG
192       lineattrs=(color=red) markerattrs=(color=red) datalabelpos=top;
193       yaxis values=(85 to 115 by 1) label='Points Per Game';
194       y2axis values=(0 to 34 by 1) label='Number of 3-Point Shot';
```

```
195        inset 'Attempts of 3-Point per game: blue star'
196        /title='Reference line represent:' titleattrs=(weight=bold) valuealign=center
197        position=topleft textattrs=(color=blue);
198        inset '   ' '    ' '3-Point Field Goals Per Game: red circle'/
199        position=topleft textattrs=(color=red);
200        run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           2.73 seconds
      user cpu time       0.09 seconds
      system cpu time     0.02 seconds
      memory              16133.28k
      OS Memory           47016.00k
      Timestamp           10/23/2019 02:12:37 AM
      Step Count                        29  Switch Count  3
      Page Faults                       0
      Page Reclaims                     5249
      Page Swaps                        0
      Voluntary Context Switches        435
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           776

NOTE: Some of the tick values have been thinned.
NOTE: Some of the tick values have been thinned.
NOTE: Some of the tick values have been thinned.
NOTE: Some of the tick values have been thinned.
NOTE: There were 9 observations read from the data set WORK.SORTSEA.

201
202        title;
203        footnote;
204
205        /*
206         * These two proc sort steps are the preparation for my next data merge step.
207         * Sort nbafinal and shoot by team name.
208         */
209        proc sort data=nbafinal out=realnbas;
210        by team;
211        run;

NOTE: There were 30 observations read from the data set WORK.NBAFINAL.
NOTE: The data set WORK.REALNBAS has 30 observations and 29 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.01 seconds
      system cpu time     0.00 seconds
      memory              1201.90k
      OS Memory           47536.00k
      Timestamp           10/23/2019 02:12:37 AM
      Step Count                        30  Switch Count  2
      Page Faults                       0
      Page Reclaims                     155
      Page Swaps                        0
      Voluntary Context Switches        17
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           272


212
213        proc sort data=shoot out=sortshoot;
214        by team;
215        run;

NOTE: There were 30 observations read from the data set WORK.SHOOT.
NOTE: The data set WORK.SORTSHOOT has 30 observations and 25 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              1201.90k
      OS Memory           47792.00k
      Timestamp           10/23/2019 02:12:37 AM
      Step Count                        31  Switch Count  2
      Page Faults                       0
      Page Reclaims                     147
      Page Swaps                        0
      Voluntary Context Switches        9
      Involuntary Context Switches      0
```

```
      Block Input Operations           0
      Block Output Operations          272


216
217         /*
218          * This data step merged realnbas and sortshoot together,
219          * and creates a new data set called bignba.
220          * Bigba has a new variables called playoff.
221          * I used if statement to separated team to two groups.
222          * I created threpp to represent the percentage of 3-point field goals
223          * by multiply 100 from the decimal representation.
224          */
225         data Bignba;
226         merge realnbas sortshoot;
227         by team;
228         threpp=threepp*100;
229         if find (team, '*') then
230         playoff='Yes';
231         else
232         playoff='No';
233         keep team age W L Winrate playoff threpp threePA threepp threep;
234         run;

NOTE: There were 30 observations read from the data set WORK.REALNBAS.
NOTE: There were 30 observations read from the data set WORK.SORTSHOOT.
NOTE: The data set WORK.BIGNBA has 30 observations and 10 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              1462.59k
      OS Memory           48048.00k
      Timestamp           10/23/2019 02:12:37 AM
      Step Count                       32  Switch Count  2
      Page Faults                      0
      Page Reclaims                    262
      Page Swaps                       0
      Voluntary Context Switches       13
      Involuntary Context Switches     0
      Block Input Operations           0
      Block Output Operations          264


235
236         /*
237          * These proc sort steps is the preparation for my next plot step.
238          * Sort bignba by winning rate, because I will use the winning rate as
239          * independent variables to make scatter plot.
240          */
241
242         proc sort data=bignba out=sortwin;
243         by descending winrate;
244         run;

NOTE: There were 30 observations read from the data set WORK.BIGNBA.
NOTE: The data set WORK.SORTWIN has 30 observations and 10 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              1193.12k
      OS Memory           48048.00k
      Timestamp           10/23/2019 02:12:37 AM
      Step Count                       33  Switch Count  2
      Page Faults                      0
      Page Reclaims                    147
      Page Swaps                       0
      Voluntary Context Switches       9
      Involuntary Context Switches     0
      Block Input Operations           0
      Block Output Operations          272


245
246         /*
247          * This proc mean step gives me the data that
248          * I will need for scatter graph's reference line.
249          * This proc step calculated the mean percentage
250          * of 3-point field goals per game by group of playoff.
```

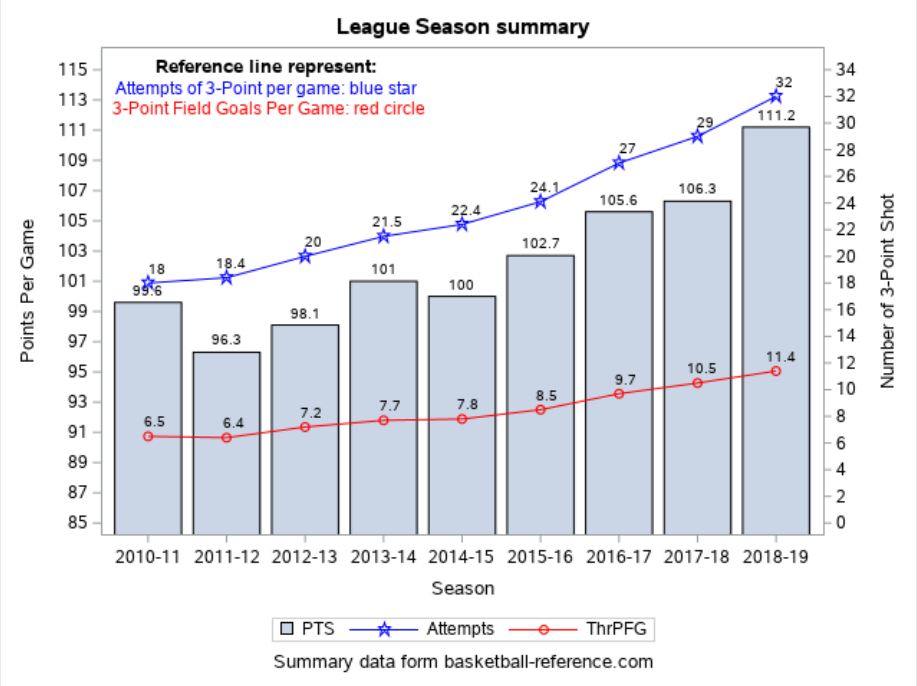
```
251         */
252        proc means data=bignba;
253        var threpp;
254        class playoff;
255        output out=meanthre;
256        run;

NOTE: There were 30 observations read from the data set WORK.BIGNBA.
NOTE: The data set WORK.MEANTHRE has 15 observations and 5 variables.
NOTE: PROCEDURE MEANS used (Total process time):
      real time           0.02 seconds
      user cpu time       0.02 seconds
      system cpu time     0.01 seconds
      memory              9592.96k
      OS Memory           57024.00k
      Timestamp           10/23/2019 02:12:37 AM
      Step Count                        34  Switch Count  3
      Page Faults                       0
      Page Reclaims                     2544
      Page Swaps                        0
      Voluntary Context Switches        36
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           272


257
258        /*
259         * These macro variables store the
260         * mean percentage of 3-point field
261         * goals per game by group of playoff
262         * and non-playoff teams.
263         */
264        %let playoff = 36.30625;
265        %let nplayoff = 34.68571;
266
267        /*
268         * This proc step creates a graph that contains scatter points
269         * for team statistic from each last season(2018-2019).
270         * Winning rate is my independent variable which is x-axis.
271         * Average percentage of 3-point field goals per game is my dependent variable which is y-axis.
272         * Green line represents the mean percentage of 3-point field goals per game for playoff teams.
273         * Black line represents the mean percentage of 3-point field goals per game for non-playoff teams.
274         */
275        proc sgplot data=sortwin;
276        title 'Team Season summary (2018-2019)';
277        footnote 'Summary data form basketball-reference.com';
278        scatter x=winrate y=threpp/ group=playoff datalabel=team;
279        yaxis label='Average 3-point field goal(%) per game';
280        refline &nplayoff / axis=y lineattrs=(color=black) label=("&nplayoff");
281        refline &playoff / axis=y lineattrs=(color=green) label=("&playoff");
282        inset 'Average 3-point field goal(%)' /title='Reference line represent:'
283        titleattrs=(weight=bold) valuealign=center position=topleft
284        textattrs=(color=blue);
285        inset '   ' '    ' ' Non-playoff team: black'/ position=topleft
286        textattrs=(color=black);
287        inset '   ' '    ' ' ' ' Playoff team: green'/ position=topleft
288        textattrs=(color=green);
289        run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.35 seconds
      user cpu time       0.06 seconds
      system cpu time     0.00 seconds
      memory              2544.06k
      OS Memory           49712.00k
      Timestamp           10/23/2019 02:12:37 AM
      Step Count                        35  Switch Count  3
      Page Faults                       0
      Page Reclaims                     498
      Page Swaps                        0
      Voluntary Context Switches        408
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           600

NOTE: There were 30 observations read from the data set WORK.SORTWIN.

290
291        /*
```

```
292         * The last date step gives a data set follows a
293         * decreasing order on winning rate. In addition,
294         * I created a Rank variable to see the ranking of
295         * the winning rate.
296         */
297        data ss;
298        set sortwin;
299        Rank=_n_;
300        run;

NOTE: There were 30 observations read from the data set WORK.SORTWIN.
NOTE: The data set WORK.SS has 30 observations and 11 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.01 seconds
      memory              946.78k
      OS Memory           49580.00k
      Timestamp           10/23/2019 02:12:37 AM
      Step Count                        36  Switch Count  2
      Page Faults                       0
      Page Reclaims                     133
      Page Swaps                        0
      Voluntary Context Switches        12
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


301
302        /*
303         * This proc print step out put the data set that
304         * contains all the variables that I used for this
305         * project. I change them to the correct format and
306         * made some label for these variables.
307         */
308        proc print data=ss noobs split='*';
309        id rank;
310        var team age W L Winrate playoff threePA threepp threep;
311        label w="Win" l='Lost' threepp='Percentage 3-point*field goals'
312        threePA='3-point*Attempts' threep="Numbers of 3-point*field goals";
313        format threepp PERCENT5.1;
314        run;

NOTE: There were 30 observations read from the data set WORK.SS.
NOTE: At least one W.D format was too small for the number to be printed. The decimal may be shifted by the "BEST" format.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.07 seconds
      user cpu time       0.08 seconds
      system cpu time     0.00 seconds
      memory              1401.93k
      OS Memory           49576.00k
      Timestamp           10/23/2019 02:12:37 AM
      Step Count                        37  Switch Count  0
      Page Faults                       0
      Page Reclaims                     259
      Page Swaps                        0
      Voluntary Context Switches        0
      Involuntary Context Switches      1
      Block Input Operations            0
      Block Output Operations           24


315
316        title;
317        footnote;
318
319
320        data work.temp;
321        text = 'DELIMIT IT';
322        post1 = find(text, 'IT','t');
323        post2 = find(text, ' IT ','t');
324        post3 = find(text, 'it');
325        post4 = find(text, 'it','i');
326        run;

NOTE: The data set WORK.TEMP has 1 observations and 5 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
```

```
system cpu time     0.00 seconds
memory              666.84k
OS Memory           49576.00k
Timestamp           10/23/2019 02:12:37 AM
Step Count                        38  Switch Count  2
Page Faults                       0
Page Reclaims                     103
Page Swaps                        0
Voluntary Context Switches        13
Involuntary Context Switches      0
Block Input Operations            0
Block Output Operations           264
```

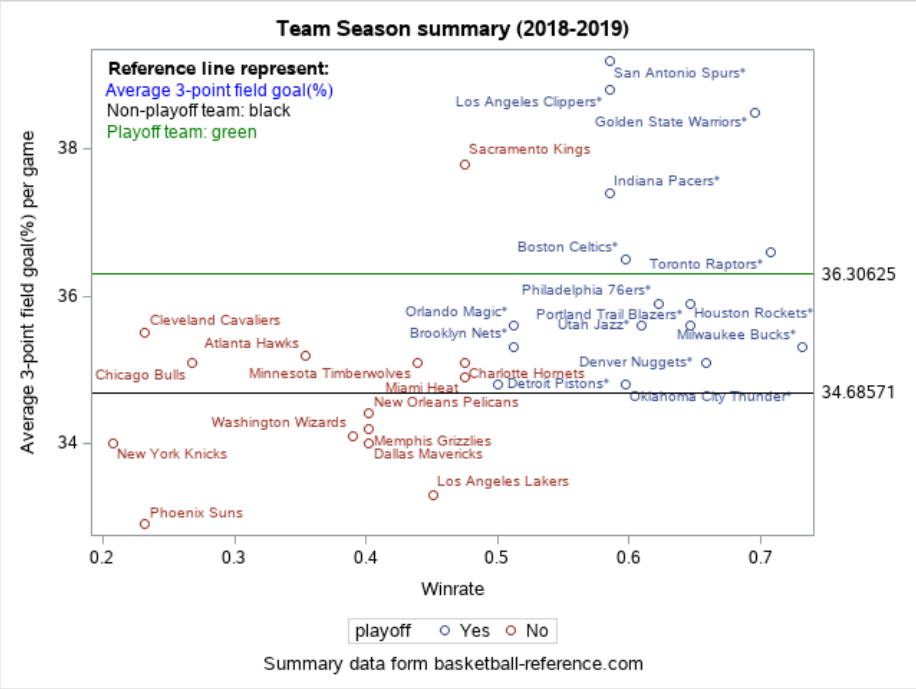
```
327
328
329
330        OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
341
```

Results: finalPro.sas



The MEANS Procedure

Analysis Variable : threpp						
playoff	N Obs	N	Mean	Std Dev	Minimum	Maximum
No	14	14	34.6857143	1.1733300	32.9000000	37.8000000
Yes	16	16	36.3062500	1.4275708	34.8000000	39.2000000



Team Season summary (2018-2019)

Rank	Team Name	Age	Win	Lost	Winrate	playoff	3-point Attempts	Percentage 3-point field goals	Numbers of 3-point field goals
1	Milwaukee Bucks*	26.9	60	22	0.73171	Yes	38.2	35%	13.5
2	Toronto Raptors*	27.3	58	24	0.70732	Yes	33.8	37%	12.4
3	Golden State Warriors*	28.4	57	25	0.69512	Yes	34.4	39%	13.3
4	Denver Nuggets*	24.9	54	28	0.65854	Yes	31.4	35%	11.0
5	Houston Rockets*	29.2	53	29	0.64634	Yes	45.4	36%	16.1
6	Portland Trail Blazers*	26.2	53	29	0.64634	Yes	30.7	36%	11.0
7	Philadelphia 76ers*	26.4	51	31	0.62195	Yes	30.2	36%	10.8
8	Utah Jazz*	27.3	50	32	0.60976	Yes	34.0	36%	12.1
9	Boston Celtics*	25.7	49	33	0.59756	Yes	34.5	37%	12.6
10	Oklahoma City Thunder*	25.7	49	33	0.59756	Yes	32.6	35%	11.4
11	Indiana Pacers*	27.0	48	34	0.58537	Yes	25.4	37%	9.5
12	Los Angeles Clippers*	27.2	48	34	0.58537	Yes	25.8	39%	10.0
13	San Antonio Spurs*	28.8	48	34	0.58537	Yes	25.3	39%	9.9
14	Brooklyn Nets*	25.4	42	40	0.51220	Yes	36.2	35%	12.8
15	Orlando Magic*	25.7	42	40	0.51220	Yes	32.1	36%	11.4
16	Detroit Pistons*	26.9	41	41	0.50000	Yes	34.8	35%	12.1
17	Charlotte Hornets	26.6	39	43	0.47561	No	33.9	35%	11.9
18	Miami Heat	27.0	39	43	0.47561	No	32.4	35%	11.3
19	Sacramento Kings	24.8	39	43	0.47561	No	29.9	38%	11.3
20	Los Angeles Lakers	26.2	37	45	0.45122	No	31.0	33%	10.3
21	Minnesota Timberwolves	26.2	36	46	0.43902	No	28.7	35%	10.1
22	Dallas Mavericks	26.9	33	49	0.40244	No	36.6	34%	12.5
23	Memphis Grizzlies	27.7	33	49	0.40244	No	28.9	34%	9.9
24	New Orleans Pelicans	25.7	33	49	0.40244	No	29.9	34%	10.3
25	Washington Wizards	26.5	32	50	0.39024	No	33.3	34%	11.3
26	Atlanta Hawks	25.1	29	53	0.35366	No	37.0	35%	13.0
27	Chicago Bulls	24.0	22	60	0.26829	No	25.9	35%	9.1
28	Cleveland Cavaliers	25.2	19	63	0.23171	No	29.1	36%	10.3
29	Phoenix Suns	24.0	19	63	0.23171	No	29.3	33%	9.6
30	New York Knicks	23.4	17	65	0.20732	No	29.5	34%	10.0

Summary data form basketball-reference.com