

Three pointer win-rate analysis

Winrate

```
winrate <- read.csv('../results/win_data/comb_winrate.csv')
year_range <- winrate$X
winrate <- subset(winrate, select=-c(X))
summary(winrate)
```

| | | | | |
|----|----------------|----------------|----------------|----------------|
| ## | BKN.NJN | NOK.NOP | LAC.SDC | VAN.MEM |
| ## | Min. :0.1460 | Min. :0.2200 | Min. :0.1460 | Min. :0.1600 |
| ## | 1st Qu.:0.3170 | 1st Qu.:0.3782 | 1st Qu.:0.2800 | 1st Qu.:0.2680 |
| ## | Median :0.4150 | Median :0.4695 | Median :0.3780 | Median :0.3410 |
| ## | Mean :0.4261 | Mean :0.4644 | Mean :0.3897 | Mean :0.4118 |
| ## | 3rd Qu.:0.5370 | 3rd Qu.:0.5580 | 3rd Qu.:0.4760 | 3rd Qu.:0.5980 |
| ## | Max. :0.6340 | Max. :0.6830 | Max. :0.6950 | Max. :0.6830 |
| ## | | NA's :23 | | NA's :16 |
| ## | WAS.WSH | CHA | OKC.SEA | KCK.SAC |
| ## | Min. :0.2320 | Min. :0.1060 | Min. :0.2440 | Min. :0.2070 |
| ## | 1st Qu.:0.3170 | 1st Qu.:0.3322 | 1st Qu.:0.4880 | 1st Qu.:0.3330 |
| ## | Median :0.4630 | Median :0.4635 | Median :0.5490 | Median :0.4020 |
| ## | Mean :0.4204 | Mean :0.4383 | Mean :0.5621 | Mean :0.4323 |
| ## | 3rd Qu.:0.5120 | 3rd Qu.:0.5370 | 3rd Qu.:0.6710 | 3rd Qu.:0.5370 |
| ## | Max. :0.5610 | Max. :0.6590 | Max. :0.7800 | Max. :0.7440 |
| ## | | NA's :11 | | |
| ## | DET | BOS | UTA | DAL |
| ## | Min. :0.1950 | Min. :0.1830 | Min. :0.2930 | Min. :0.1340 |
| ## | 1st Qu.:0.3900 | 1st Qu.:0.4390 | 1st Qu.:0.5000 | 1st Qu.:0.3703 |
| ## | Median :0.5610 | Median :0.5850 | Median :0.5730 | Median :0.5370 |
| ## | Mean :0.5181 | Mean :0.5669 | Mean :0.5558 | Mean :0.5091 |
| ## | 3rd Qu.:0.6340 | 3rd Qu.:0.6950 | 3rd Qu.:0.6590 | 3rd Qu.:0.6460 |
| ## | Max. :0.7800 | Max. :0.8170 | Max. :0.7800 | Max. :0.8170 |
| ## | | | | NA's :1 |
| ## | MIA | IND | PHI | CHI |
| ## | Min. :0.1830 | Min. :0.2440 | Min. :0.122 | Min. :0.1830 |
| ## | 1st Qu.:0.4390 | 1st Qu.:0.4390 | 1st Qu.:0.378 | 1st Qu.:0.3660 |
| ## | Median :0.5305 | Median :0.5000 | Median :0.500 | Median :0.5490 |
| ## | Mean :0.5236 | Mean :0.5052 | Mean :0.487 | Mean :0.5281 |
| ## | 3rd Qu.:0.6593 | 3rd Qu.:0.5850 | 3rd Qu.:0.598 | 3rd Qu.:0.6710 |
| ## | Max. :0.8050 | Max. :0.7440 | Max. :0.793 | Max. :0.8780 |
| ## | NA's :9 | | | |
| ## | POR | LAL | ATL | CLE |
| ## | Min. :0.2560 | Min. :0.2070 | Min. :0.1590 | Min. :0.1830 |
| ## | 1st Qu.:0.5000 | 1st Qu.:0.5490 | 1st Qu.:0.4270 | 1st Qu.:0.3540 |
| ## | Median :0.5730 | Median :0.6830 | Median :0.5240 | Median :0.4510 |
| ## | Mean :0.5598 | Mean :0.6265 | Mean :0.5123 | Mean :0.4756 |
| ## | 3rd Qu.:0.6220 | 3rd Qu.:0.7070 | 3rd Qu.:0.6100 | 3rd Qu.:0.5730 |
| ## | Max. :0.7680 | Max. :0.8170 | Max. :0.7320 | Max. :0.8050 |
| ## | | | | |
| ## | DEN | HOU | SAS | PHX |
| ## | Min. :0.1340 | Min. :0.1710 | Min. :0.2440 | Min. :0.2800 |
| ## | 1st Qu.:0.4020 | 1st Qu.:0.5000 | 1st Qu.:0.5850 | 1st Qu.:0.4880 |
| ## | Median :0.5000 | Median :0.5490 | Median :0.6710 | Median :0.5850 |

```
## Mean :0.4734 Mean :0.5462 Mean :0.6255 Mean :0.5671
## 3rd Qu.:0.5730 3rd Qu.:0.6220 3rd Qu.:0.7200 3rd Qu.:0.6710
## Max. :0.6950 Max. :0.7070 Max. :0.8170 Max. :0.7560
##
## ORL MIL GSW MIN
## Min. :0.2200 Min. :0.1830 Min. :0.2070 Min. :0.1830
## 1st Qu.:0.4025 1st Qu.:0.4020 1st Qu.:0.3480 1st Qu.:0.2620
## Median :0.5000 Median :0.5000 Median :0.4390 Median :0.3780
## Mean :0.4933 Mean :0.4909 Mean :0.4433 Mean :0.3927
## 3rd Qu.:0.6220 3rd Qu.:0.5980 3rd Qu.:0.5240 3rd Qu.:0.5185
## Max. :0.7320 Max. :0.7320 Max. :0.8900 Max. :0.7070
## NA's :10 NA's :10
## TOR NYK
## Min. :0.195 Min. :0.2070
## 1st Qu.:0.348 1st Qu.:0.3900
## Median :0.415 Median :0.4760
## Mean :0.438 Mean :0.4864
## 3rd Qu.:0.549 3rd Qu.:0.5850
## Max. :0.683 Max. :0.7320
## NA's :16
```

Threepointer percentage data

```
percent3p <- read.csv('.././results/team_data/comb_3pper.csv')
year_range <- percent3p$X
percent3p <- subset(percent3p, select=-c(X))
summary(percent3p)
```

```
## BKN.NJN NOK.NOP LAC.SDC VAN.MEM
## Min. :0.2010 Min. :0.3150 Min. :0.188 Min. :0.3070
## 1st Qu.:0.3010 1st Qu.:0.3443 1st Qu.:0.289 1st Qu.:0.3340
## Median :0.3310 Median :0.3625 Median :0.326 Median :0.3450
## Mean :0.3138 Mean :0.3561 Mean :0.314 Mean :0.3455
## 3rd Qu.:0.3430 3rd Qu.:0.3685 3rd Qu.:0.352 3rd Qu.:0.3600
## Max. :0.3760 Max. :0.3890 Max. :0.376 Max. :0.3740
## NA's :23 NA's :16
## WAS.WSH CHA OKC.SEA KCK.SAC
## Min. :0.1880 Min. :0.2950 Min. :0.1930 Min. :0.2000
## 1st Qu.:0.2740 1st Qu.:0.3290 1st Qu.:0.3330 1st Qu.:0.3160
## Median :0.3240 Median :0.3470 Median :0.3470 Median :0.3490
## Mean :0.3102 Mean :0.3495 Mean :0.3354 Mean :0.3303
## 3rd Qu.:0.3530 3rd Qu.:0.3658 3rd Qu.:0.3640 3rd Qu.:0.3670
## Max. :0.4070 Max. :0.4280 Max. :0.3990 Max. :0.4010
## NA's :11
## UTA NYK BOS GSW
## Min. :0.2070 Min. :0.2200 Min. :0.2100 Min. :0.2230
## 1st Qu.:0.3190 1st Qu.:0.3320 1st Qu.:0.3150 1st Qu.:0.2930
## Median :0.3360 Median :0.3460 Median :0.3490 Median :0.3370
## Mean :0.3318 Mean :0.3347 Mean :0.3336 Mean :0.3303
## 3rd Qu.:0.3610 3rd Qu.:0.3620 3rd Qu.:0.3650 3rd Qu.:0.3560
## Max. :0.3850 Max. :0.3830 Max. :0.3970 Max. :0.4160
##
## TOR CHI ATL DEN
## Min. :0.3160 Min. :0.1710 Min. :0.1220 Min. :0.1900
## 1st Qu.:0.3430 1st Qu.:0.2890 1st Qu.:0.3060 1st Qu.:0.3020
```

```
## Median :0.3630 Median :0.3480 Median :0.3320 Median :0.3320
## Mean :0.3587 Mean :0.3238 Mean :0.3156 Mean :0.3196
## 3rd Qu.:0.3710 3rd Qu.:0.3660 3rd Qu.:0.3560 3rd Qu.:0.3430
## Max. :0.3920 Max. :0.4030 Max. :0.3800 Max. :0.3880
## NA's :16
## HOU MIN DAL MIL
## Min. :0.1780 Min. :0.2480 Min. :0.2280 Min. :0.2190
## 1st Qu.:0.3120 1st Qu.:0.3230 1st Qu.:0.3220 1st Qu.:0.3240
## Median :0.3460 Median :0.3390 Median :0.3460 Median :0.3450
## Mean :0.3248 Mean :0.3334 Mean :0.3391 Mean :0.3377
## 3rd Qu.:0.3620 3rd Qu.:0.3515 3rd Qu.:0.3668 3rd Qu.:0.3630
## Max. :0.3750 Max. :0.3780 Max. :0.3910 Max. :0.3830
## NA's :10 NA's :1
## MIA PHI PHX SAS
## Min. :0.2930 Min. :0.2160 Min. :0.2080 Min. :0.176
## 1st Qu.:0.3410 1st Qu.:0.2990 1st Qu.:0.3190 1st Qu.:0.292
## Median :0.3545 Median :0.3200 Median :0.3430 Median :0.354
## Mean :0.3501 Mean :0.3135 Mean :0.3349 Mean :0.332
## 3rd Qu.:0.3640 3rd Qu.:0.3430 3rd Qu.:0.3720 3rd Qu.:0.375
## Max. :0.3960 Max. :0.3790 Max. :0.4120 Max. :0.407
## NA's :9
## DET CLE ORL IND
## Min. :0.1550 Min. :0.182 Min. :0.2950 Min. :0.1610
## 1st Qu.:0.3000 1st Qu.:0.332 1st Qu.:0.3425 1st Qu.:0.3280
## Median :0.3440 Median :0.352 Median :0.3560 Median :0.3480
## Mean :0.3217 Mean :0.337 Mean :0.3534 Mean :0.3305
## 3rd Qu.:0.3560 3rd Qu.:0.373 3rd Qu.:0.3715 3rd Qu.:0.3680
## Max. :0.4040 Max. :0.407 Max. :0.3860 Max. :0.3920
## NA's :10
## POR LAL
## Min. :0.1690 Min. :0.1040
## 1st Qu.:0.3090 1st Qu.:0.3040
## Median :0.3460 Median :0.3440
## Mean :0.3242 Mean :0.3178
## 3rd Qu.:0.3610 3rd Qu.:0.3540
## Max. :0.3830 Max. :0.3810
##
```

Threepointer attempts data

```
attempts3p <- read.csv('../results/team_data/comb_3pattempt.csv')
year_range <- attempts3p$X
attempts3p <- subset(attempts3p, select=-c(X))
summary(attempts3p)
```

```
## BKN.NJN NOK.NOP LAC.SDC VAN.MEM
## Min. : 1.70 Min. :10.80 Min. : 1.60 Min. : 9.10
## 1st Qu.: 6.00 1st Qu.:15.07 1st Qu.: 5.00 1st Qu.:12.40
## Median :13.20 Median :17.00 Median :12.50 Median :13.80
## Mean :11.78 Mean :16.92 Mean :11.67 Mean :14.47
## 3rd Qu.:16.80 3rd Qu.:19.27 3rd Qu.:16.80 3rd Qu.:16.00
## Max. :23.40 Max. :23.80 Max. :26.90 Max. :21.70
## NA's :23 NA's :16
## WAS.WSH CHA OKC.SEA KCK.SAC
## Min. : 1.70 Min. : 4.50 Min. : 1.40 Min. : 1.40
```

| | | | | |
|----|----------------|---------------|----------------|---------------|
| ## | 1st Qu.: 3.50 | 1st Qu.:11.05 | 1st Qu.: 7.40 | 1st Qu.: 7.00 |
| ## | Median :10.90 | Median :14.70 | Median :15.00 | Median :14.10 |
| ## | Mean :10.74 | Mean :13.77 | Mean :12.98 | Mean :12.09 |
| ## | 3rd Qu.:15.50 | 3rd Qu.:17.05 | 3rd Qu.:19.40 | 3rd Qu.:17.20 |
| ## | Max. :24.20 | Max. :29.40 | Max. :23.70 | Max. :22.40 |
| ## | | NA's :11 | | |
| ## | ORL | POR | CHI | NYK |
| ## | Min. : 4.80 | Min. : 1.60 | Min. : 1.4 | Min. : 1.60 |
| ## | 1st Qu.:11.15 | 1st Qu.: 6.90 | 1st Qu.: 5.2 | 1st Qu.: 6.90 |
| ## | Median :17.20 | Median :13.40 | Median :12.2 | Median :13.60 |
| ## | Mean :16.81 | Mean :12.55 | Mean :11.0 | Mean :13.27 |
| ## | 3rd Qu.:20.25 | 3rd Qu.:16.90 | 3rd Qu.:15.9 | 3rd Qu.:17.70 |
| ## | Max. :27.30 | Max. :28.50 | Max. :22.3 | Max. :28.90 |
| ## | NA's :10 | | | |
| ## | DET | SAS | DEN | GSW |
| ## | Min. : 1.00 | Min. : 1.00 | Min. : 1.50 | Min. : 1.50 |
| ## | 1st Qu.: 4.90 | 1st Qu.: 3.80 | 1st Qu.: 5.50 | 1st Qu.: 7.70 |
| ## | Median :13.20 | Median :13.30 | Median :12.90 | Median :12.10 |
| ## | Mean :11.59 | Mean :11.66 | Mean :12.36 | Mean :13.57 |
| ## | 3rd Qu.:16.50 | 3rd Qu.:18.50 | 3rd Qu.:18.40 | 3rd Qu.:20.50 |
| ## | Max. :26.20 | Max. :22.50 | Max. :24.80 | Max. :31.60 |
| ## | | | | |
| ## | LAL | MIL | PHX | ATL |
| ## | Min. : 1.10 | Min. : 1.60 | Min. : 1.90 | Min. : 0.90 |
| ## | 1st Qu.: 5.80 | 1st Qu.: 7.00 | 1st Qu.: 5.30 | 1st Qu.: 5.00 |
| ## | Median :15.50 | Median :12.40 | Median :14.00 | Median :12.70 |
| ## | Mean :13.16 | Mean :12.06 | Mean :13.41 | Mean :12.27 |
| ## | 3rd Qu.:18.50 | 3rd Qu.:17.20 | 3rd Qu.:19.60 | 3rd Qu.:17.70 |
| ## | Max. :24.80 | Max. :22.10 | Max. :25.80 | Max. :28.40 |
| ## | | | | |
| ## | IND | MIA | BOS | MIN |
| ## | Min. : 1.70 | Min. : 3.60 | Min. : 2.20 | Min. : 3.60 |
| ## | 1st Qu.: 6.10 | 1st Qu.:13.10 | 1st Qu.: 4.80 | 1st Qu.: 9.25 |
| ## | Median :13.60 | Median :16.60 | Median :15.20 | Median :12.40 |
| ## | Mean :12.61 | Mean :15.36 | Mean :12.58 | Mean :12.36 |
| ## | 3rd Qu.:18.70 | 3rd Qu.:18.20 | 3rd Qu.:17.50 | 3rd Qu.:15.15 |
| ## | Max. :24.60 | Max. :22.70 | Max. :26.30 | Max. :21.60 |
| ## | | NA's :9 | | NA's :10 |
| ## | DAL | CLE | UTA | HOU |
| ## | Min. : 2.000 | Min. : 1.50 | Min. : 1.200 | Min. : 1.40 |
| ## | 1st Qu.: 7.825 | 1st Qu.: 5.80 | 1st Qu.: 5.000 | 1st Qu.: 6.00 |
| ## | Median :15.050 | Median :10.70 | Median : 9.300 | Median :17.20 |
| ## | Mean :13.878 | Mean :11.78 | Mean : 9.349 | Mean :15.34 |
| ## | 3rd Qu.:19.900 | 3rd Qu.:18.20 | 3rd Qu.:12.800 | 3rd Qu.:21.40 |
| ## | Max. :28.600 | Max. :29.60 | Max. :23.900 | Max. :32.70 |
| ## | NA's :1 | | | |
| ## | PHI | TOR | | |
| ## | Min. : 1.00 | Min. :13.2 | | |
| ## | 1st Qu.: 6.60 | 1st Qu.:14.2 | | |
| ## | Median : 9.90 | Median :16.3 | | |
| ## | Mean :10.58 | Mean :17.3 | | |
| ## | 3rd Qu.:14.60 | 3rd Qu.:19.8 | | |
| ## | Max. :27.50 | Max. :25.1 | | |
| ## | | NA's :16 | | |

Threepointer made data

```
made3p <- read.csv('.../results/team_data/comb_3pmade.csv')
year_range <- made3p$X
made3p <- subset(made3p, select=-c(X))
summary(made3p)
```

```
##      BKN.NJN      NOK.NOP      LAC.SDC      VAN.MEM
## Min.    :0.400   Min.    :3.700   Min.    : 0.300   Min.    :3.000
## 1st Qu.:1.800   1st Qu.:5.175   1st Qu.: 1.400   1st Qu.:4.100
## Median :4.400   Median :6.200   Median : 4.000   Median :4.900
## Mean    :3.965   Mean    :6.043   Mean    : 3.941   Mean    :5.029
## 3rd Qu.:5.800   3rd Qu.:6.950   3rd Qu.: 5.900   3rd Qu.:5.700
## Max.    :8.600   Max.    :8.600   Max.    :10.100   Max.    :7.600
##      NA's      :23      NA's      :16
##      WAS.WSH      CHA      OKC.SEA      KCK.SAC
## Min.    :0.400   Min.    : 1.400   Min.    :0.300   Min.    :0.300
## 1st Qu.:0.900   1st Qu.: 4.025   1st Qu.:2.500   1st Qu.:2.600
## Median :3.700   Median : 5.000   Median :5.600   Median :5.200
## Mean    :3.624   Mean    : 4.873   Mean    :4.619   Mean    :4.216
## 3rd Qu.:5.300   3rd Qu.: 6.075   3rd Qu.:7.100   3rd Qu.:6.000
## Max.    :8.600   Max.    :10.600   Max.    :8.800   Max.    :8.000
##      NA's      :11
##      GSW      SAS      BOS      UTA
## Min.    : 0.3   Min.    :0.20   Min.    :0.500   Min.    :0.300
## 1st Qu.: 2.3   1st Qu.:1.10   1st Qu.:1.300   1st Qu.:1.600
## Median : 3.9   Median :4.60   Median :5.100   Median :3.000
## Mean    : 4.8   Mean    :4.23   Mean    :4.376   Mean    :3.235
## 3rd Qu.: 7.6   3rd Qu.:6.80   3rd Qu.:6.600   3rd Qu.:4.300
## Max.    :13.1   Max.    :8.50   Max.    :8.800   Max.    :8.500
##
##      POR      NYK      HOU      MIA
## Min.    : 0.300   Min.    : 0.400   Min.    : 0.300   Min.    :1.100
## 1st Qu.: 2.300   1st Qu.: 2.300   1st Qu.: 1.900   1st Qu.:4.175
## Median : 4.600   Median : 4.700   Median : 6.100   Median :5.850
## Mean    : 4.392   Mean    : 4.673   Mean    : 5.359   Mean    :5.450
## 3rd Qu.: 6.000   3rd Qu.: 6.500   3rd Qu.: 7.800   3rd Qu.:6.700
## Max.    :10.500   Max.    :10.900   Max.    :11.400   Max.    :8.700
##      NA's      :9
##      DET      CHI      MIL      TOR
## Min.    :0.200   Min.    :0.200   Min.    :0.400   Min.    :4.200
## 1st Qu.:1.500   1st Qu.:1.700   1st Qu.:2.300   1st Qu.:5.000
## Median :4.600   Median :4.000   Median :4.500   Median :5.800
## Mean    :4.024   Mean    :3.849   Mean    :4.243   Mean    :6.224
## 3rd Qu.:6.000   3rd Qu.:6.000   3rd Qu.:6.200   3rd Qu.:7.200
## Max.    :9.000   Max.    :7.900   Max.    :7.900   Max.    :8.900
##      NA's      :16
##      CLE      PHI      DEN      DAL
## Min.    : 0.300   Min.    :0.200   Min.    :0.300   Min.    :0.500
## 1st Qu.: 2.000   1st Qu.:2.000   1st Qu.:1.700   1st Qu.:2.550
## Median : 3.600   Median :3.200   Median :4.000   Median :5.150
## Mean    : 4.192   Mean    :3.503   Mean    :4.165   Mean    :4.908
## 3rd Qu.: 6.200   3rd Qu.:5.300   3rd Qu.:6.400   3rd Qu.:7.100
## Max.    :10.700   Max.    :9.300   Max.    :8.600   Max.    :9.800
##      NA's      :1
```

```
##          IND          ATL          LAL          MIN
## Min.    :0.300   Min.    : 0.100   Min.    :0.10   Min.    :0.900
## 1st Qu.:2.200   1st Qu.: 1.500   1st Qu.:2.00   1st Qu.:3.200
## Median :4.800   Median : 4.100   Median :5.40   Median :4.000
## Mean    :4.449   Mean    : 4.241   Mean    :4.53   Mean    :4.185
## 3rd Qu.:6.600   3rd Qu.: 6.400   3rd Qu.:6.50   3rd Qu.:5.200
## Max.    :9.200   Max.    :10.000   Max.    :9.40   Max.    :7.300
##
##                                     NA's    :10
##          PHX          ORL
## Min.    : 0.400   Min.    : 1.400
## 1st Qu.: 1.700   1st Qu.: 3.900
## Median : 4.900   Median : 6.200
## Mean    : 4.814   Mean    : 6.044
## 3rd Qu.: 7.100   3rd Qu.: 7.600
## Max.    :10.200   Max.    :10.300
##
##                                     NA's    :10
```

Misc data

```
eff_fg_per <- read.csv('../results/misc_data/comb_Effective_Field_Goal_Percentage.csv')
ft_per_fg <- read.csv('../results/misc_data/comb_Free_Throws_Per_Field_Goal_Attempt.csv')
off_reb <- read.csv('../results/misc_data/comb_Offensive_Rebound_Percentage.csv')
turn_per <- read.csv('../results/misc_data/comb_Turnover_Percentage.csv')
year_range <- eff_fg_per$X
eff_fg_per <- subset(eff_fg_per, select=-c(X))
ft_per_fg <- subset(ft_per_fg, select=-c(X))
off_reb <- subset(off_reb, select=-c(X))
turn_per <- subset(turn_per, select=-c(X))
```

Correlation between winrate and threepoint percentage in the last decades

```
get_lmmodel <- function(start_yr, end_yr, ind_var1, ind_var2, ind_var3, ind_var4)
{
  winrate.list <- c()
  indvar1.list <- c()
  indvar2.list <- c()
  indvar3.list <- c()
  indvar4.list <- c()
  xx.list <- setNames(split(ind_var1, seq(nrow(ind_var1))), year_range)
  yy.list <- setNames(split(ind_var2, seq(nrow(ind_var2))), year_range)
  aa.list <- setNames(split(ind_var3, seq(nrow(ind_var3))), year_range)
  bb.list <- setNames(split(ind_var4, seq(nrow(ind_var4))), year_range)
  yz.list <- setNames(split(winrate, seq(nrow(winrate))), year_range)
  # Change the year range to get models for different time ranges
  yrs <- seq(match(start_yr, year_range), match(end_yr, year_range))
  for (yr_ind in yrs)
  {
    indvar1.list <- c(indvar1.list, as.numeric(xx.list[[yr_ind]]))
    indvar2.list <- c(indvar2.list, as.numeric(yy.list[[yr_ind]]))
    indvar3.list <- c(indvar3.list, as.numeric(aa.list[[yr_ind]]))
    indvar4.list <- c(indvar4.list, as.numeric(bb.list[[yr_ind]]))
    winrate.list <- c(winrate.list, as.numeric(yz.list[[yr_ind]]))
  }
  # Check if these are in order
  # model <- lm(winrate.list ~ indvar1.list*indvar2.list)
  model <- lm(winrate.list ~ indvar1.list+indvar2.list+indvar3.list+indvar4.list)
```

```

}
model <- get_lmmodel(2005, 2016, ft_per_fg, eff_fg_per, off_reb, turn_per)
summary(model)

```

```

##
## Call:
## lm(formula = winrate.list ~ indvar1.list + indvar2.list + indvar3.list +
##     indvar4.list)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.35369 -0.11755  0.00681  0.11173  0.32297
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.076508   0.253879  -0.301  0.763320
## indvar1.list   0.498988   0.288081   1.732  0.084125 .
## indvar2.list   1.378569   0.413432   3.334  0.000945 ***
## indvar3.list  -0.001190   0.003083  -0.386  0.699671
## indvar4.list  -0.013911   0.008482  -1.640  0.101874
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1505 on 354 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.04203,    Adjusted R-squared:  0.0312
## F-statistic: 3.883 on 4 and 354 DF,  p-value: 0.004224

```

```
AIC(model)
```

```
## [1] -333.9732
```

Random forest

```
#set.seed(415)
```

```

get_randforest <- function(start_yr, end_yr, ind_var1, ind_var2, ind_var3, ind_var4, ind_var5, ind_var6)
{
  winrate.list <- c()
  indvar1.list <- c()
  indvar2.list <- c()
  indvar3.list <- c()
  indvar4.list <- c()
  indvar5.list <- c()
  indvar6.list <- c()
  xx.list <- setNames(split(ind_var1, seq(nrow(ind_var1))), year_range)
  yy.list <- setNames(split(ind_var2, seq(nrow(ind_var2))), year_range)
  aa.list <- setNames(split(ind_var3, seq(nrow(ind_var3))), year_range)
  bb.list <- setNames(split(ind_var4, seq(nrow(ind_var4))), year_range)
  cc.list <- setNames(split(ind_var5, seq(nrow(ind_var5))), year_range)
  dd.list <- setNames(split(ind_var6, seq(nrow(ind_var6))), year_range)
  yz.list <- setNames(split(winrate, seq(nrow(winrate))), year_range)
  # Change the year range to get models for different time ranges
  yrs <- seq(match(start_yr, year_range), match(end_yr, year_range))
  for (yr_ind in yrs)
  {

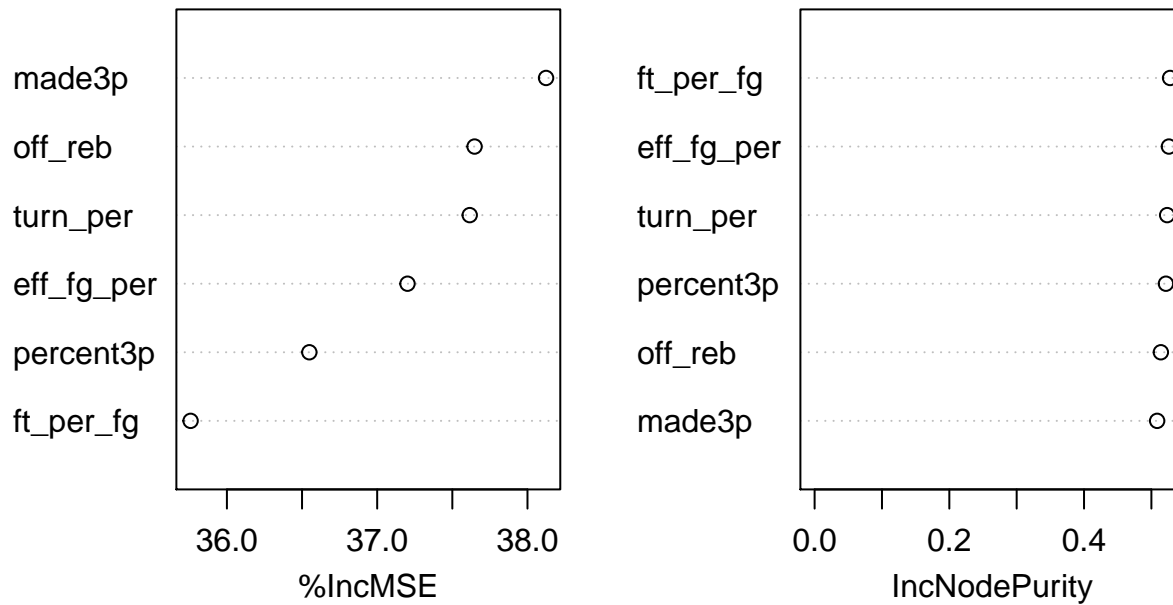
```

```

indvar1.list <- c(indvar1.list, as.numeric(xx.list[[yr_ind]]))
indvar2.list <- c(indvar2.list, as.numeric(yy.list[[yr_ind]]))
indvar3.list <- c(indvar3.list, as.numeric(aa.list[[yr_ind]]))
indvar4.list <- c(indvar4.list, as.numeric(bb.list[[yr_ind]]))
indvar5.list <- c(indvar5.list, as.numeric(cc.list[[yr_ind]]))
indvar6.list <- c(indvar6.list, as.numeric(dd.list[[yr_ind]]))
winrate.list <- c(winrate.list, as.numeric(yz.list[[yr_ind]]))
}
smp_size <- floor(0.8 * length(winrate.list))
train_ind <- sample(seq_len(length(winrate.list)), size=smp_size)
#Training set
ft_per_fg <- indvar1.list[train_ind]
eff_fg_per <- indvar1.list[train_ind]
off_reb <- indvar1.list[train_ind]
turn_per <- indvar1.list[train_ind]
made3p <- indvar1.list[train_ind]
percent3p <- indvar1.list[train_ind]
winrate <- winrate.list[train_ind]
train_data <- data.frame(ft_per_fg=ft_per_fg, eff_fg_per=eff_fg_per, off_reb=off_reb, turn_per=turn_per,
# Testing set
ft_per_fg <- indvar1.list[-train_ind]
eff_fg_per <- indvar1.list[-train_ind]
off_reb <- indvar1.list[-train_ind]
turn_per <- indvar1.list[-train_ind]
made3p <- indvar1.list[-train_ind]
percent3p <- indvar1.list[-train_ind]
winrate <- winrate.list[-train_ind]
test_data <- data.frame(ft_per_fg=ft_per_fg, eff_fg_per=eff_fg_per, off_reb=off_reb, turn_per=turn_per,
# Check if these are in order
model <- randomForest(winrate~ft_per_fg+eff_fg_per+off_reb+turn_per+made3p+percent3p, data=train_data)
prediction <- predict(model, test_data, OOB=TRUE)
pred_score = mean((test_data$winrate - prediction)^2)
return(list('model'=model, 'train_data'=train_data, 'prediction'=pred_score))
}
forest_stuff <- get_randforest(2006, 2016, ft_per_fg, eff_fg_per, off_reb, turn_per, made3p, percent3p)
varImpPlot(forest_stuff$model)

```


forest_stuff\$model



```
print(forest_stuff$prediction)
```

```
## [1] 0.03260088
```

Don't really need to do CV since randomForest already partitions randomly

```
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```
##
```

```
## Attaching package: 'ggplot2'
```

```
## The following object is masked from 'package:randomForest':
```

```
##
```

```
## margin
```

```
rf <- forest_stuff$model
```

```
imp <- importance(rf)
```

```
imp = data.frame(type=rownames(imp), importance(rf), check.names=F)
```

```
imp$type = reorder(imp$type, imp$`%IncMSE`)
```

```
ggplot(data=imp, aes(x=type, y=`%IncMSE`)) + geom_bar(stat='identity') + geom_hline(yintercept=abs(min(
```

