Kobe\_Analysis

Pankaj Kumar

7/14/2019

## Some observation on the data

length(kobe\_data)

## [1] 29

nrow(kobe\_data)

## [1] 20697

ncol(kobe\_data)

## [1] 29

dim(kobe\_data)

## [1] 20697 29

## check if there are any null or empty values are there

## recId action\_type combined\_shot\_type   
## FALSE FALSE FALSE   
## game\_event\_id game\_id lat   
## FALSE FALSE FALSE   
## loc\_x loc\_y lon   
## FALSE FALSE FALSE   
## minutes\_remaining period playoffs   
## FALSE FALSE FALSE   
## season seconds\_remaining shot\_distance   
## FALSE FALSE FALSE   
## shot\_made\_flag shot\_type shot\_zone\_area   
## FALSE FALSE FALSE   
## shot\_zone\_basic shot\_zone\_range team\_id   
## FALSE FALSE FALSE   
## team\_name game\_date matchup   
## FALSE FALSE FALSE   
## opponent shot\_id attendance   
## FALSE FALSE FALSE   
## arena\_temp avgnoisedb   
## FALSE FALSE

### Not null or missing

## recId action\_type combined\_shot\_type   
## FALSE FALSE FALSE   
## game\_event\_id game\_id lat   
## FALSE FALSE FALSE   
## loc\_x loc\_y lon   
## FALSE FALSE FALSE   
## minutes\_remaining period playoffs   
## FALSE FALSE FALSE   
## season seconds\_remaining shot\_distance   
## FALSE FALSE FALSE   
## shot\_made\_flag shot\_type shot\_zone\_area   
## FALSE FALSE FALSE   
## shot\_zone\_basic shot\_zone\_range team\_id   
## FALSE FALSE FALSE   
## team\_name game\_date matchup   
## FALSE FALSE FALSE   
## opponent shot\_id attendance   
## FALSE FALSE FALSE   
## arena\_temp avgnoisedb   
## FALSE FALSE

### Transformations :

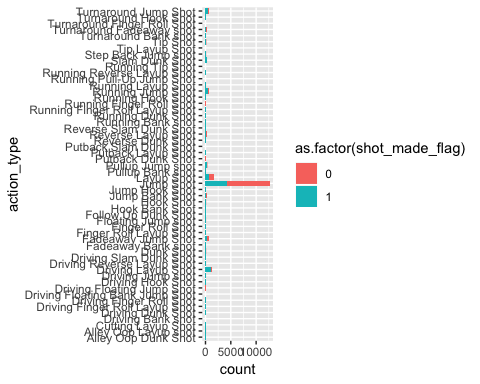
minutes remaining and second remaining can be compined to make a new variable called time remaining

# plots

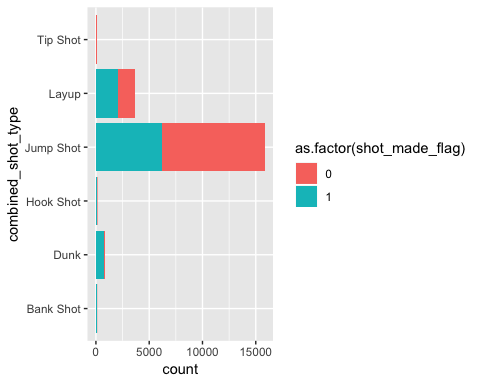
## Factors

Graphs of all the variable with alphanumeric value.

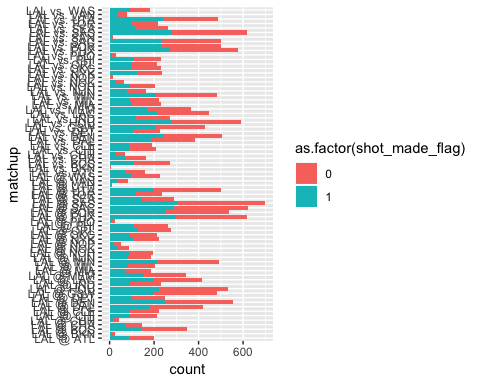
### action\_type



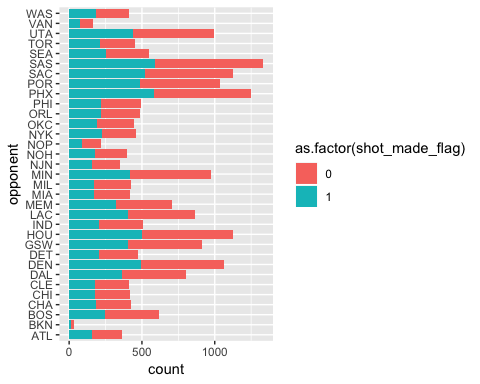
### combined\_shot\_type



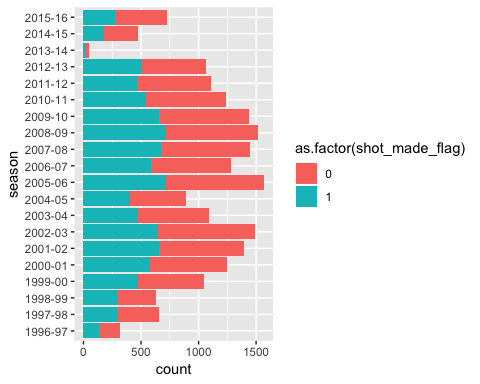
### matchup



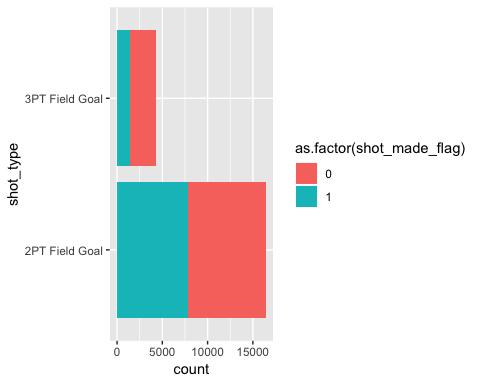
### opponent



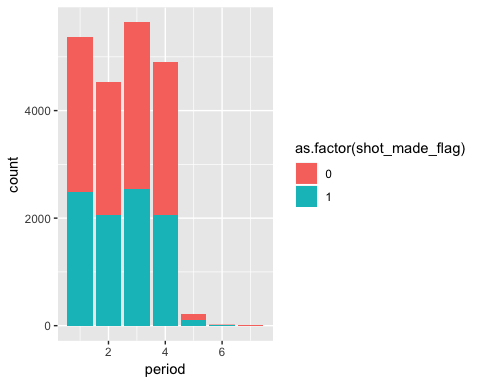
### Season



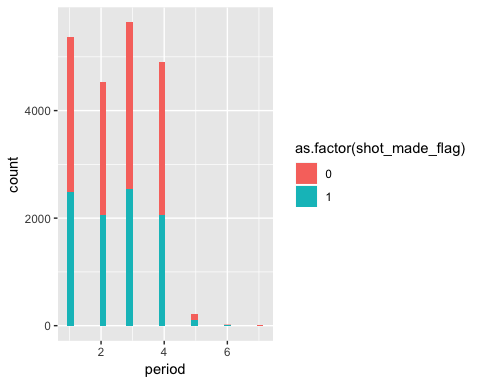
### shot\_type



### period

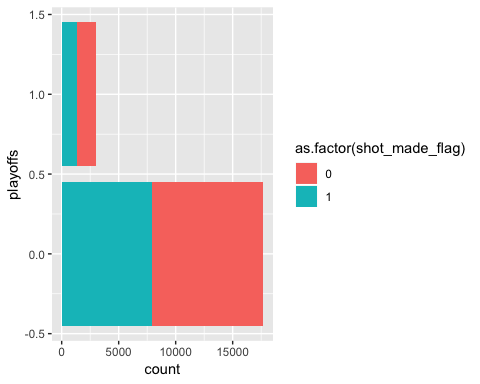


## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

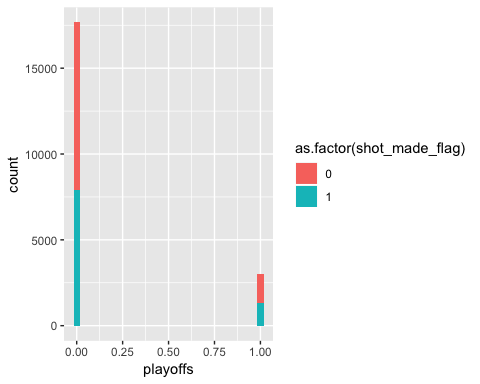


## <ggproto object: Class CoordFlip, CoordCartesian, Coord, gg>  
## aspect: function  
## backtransform\_range: function  
## clip: on  
## default: FALSE  
## distance: function  
## expand: TRUE  
## is\_free: function  
## is\_linear: function  
## labels: function  
## limits: list  
## modify\_scales: function  
## range: function  
## render\_axis\_h: function  
## render\_axis\_v: function  
## render\_bg: function  
## render\_fg: function  
## setup\_data: function  
## setup\_layout: function  
## setup\_panel\_params: function  
## setup\_params: function  
## transform: function  
## super: <ggproto object: Class CoordFlip, CoordCartesian, Coord, gg>

### playoffs

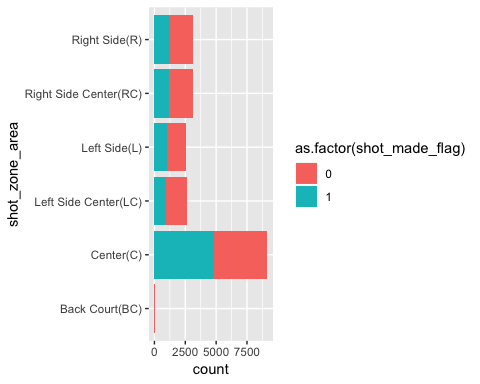


## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

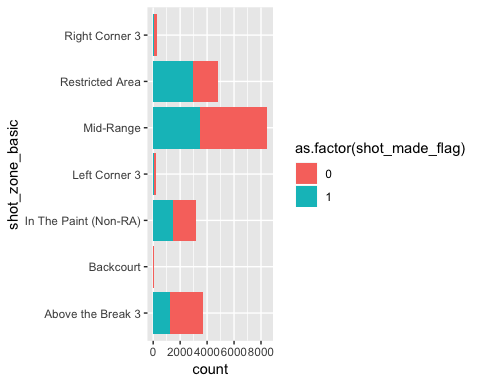


## <ggproto object: Class CoordFlip, CoordCartesian, Coord, gg>  
## aspect: function  
## backtransform\_range: function  
## clip: on  
## default: FALSE  
## distance: function  
## expand: TRUE  
## is\_free: function  
## is\_linear: function  
## labels: function  
## limits: list  
## modify\_scales: function  
## range: function  
## render\_axis\_h: function  
## render\_axis\_v: function  
## render\_bg: function  
## render\_fg: function  
## setup\_data: function  
## setup\_layout: function  
## setup\_panel\_params: function  
## setup\_params: function  
## transform: function  
## super: <ggproto object: Class CoordFlip, CoordCartesian, Coord, gg>

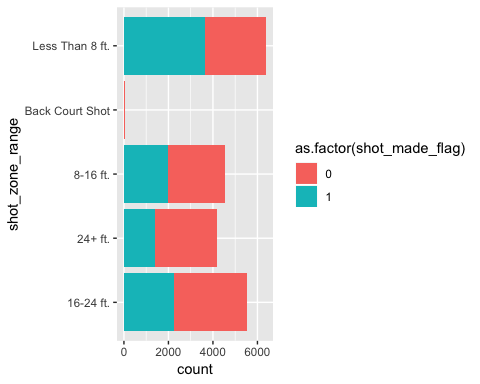
### shot\_zone\_area



### shot\_zone\_basic



### shot\_zone\_range

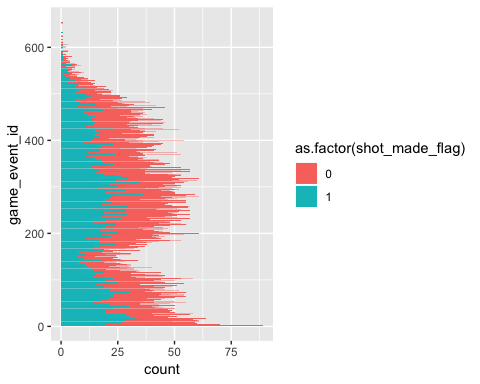


## Numerical variables that can be treated as factors

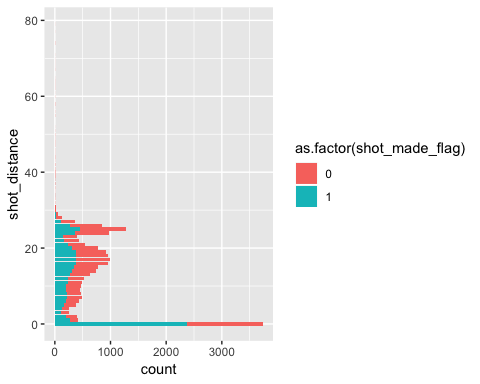
### game\_event\_id

## Continuous var

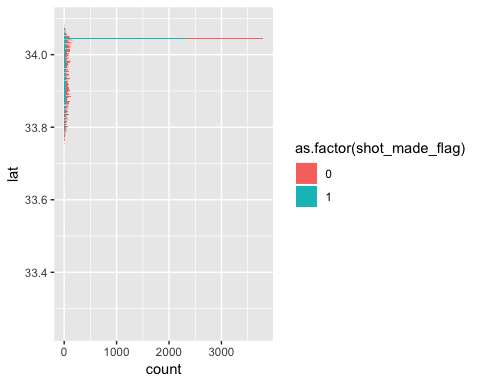
### game\_event\_id



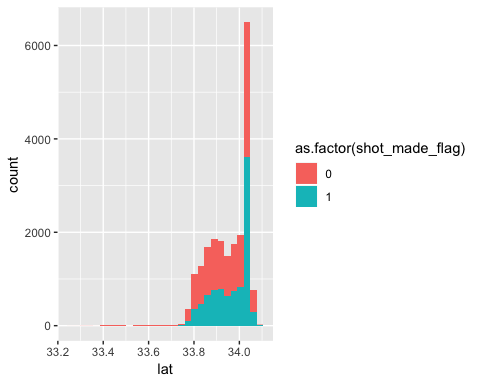
### shot\_distance



### lat

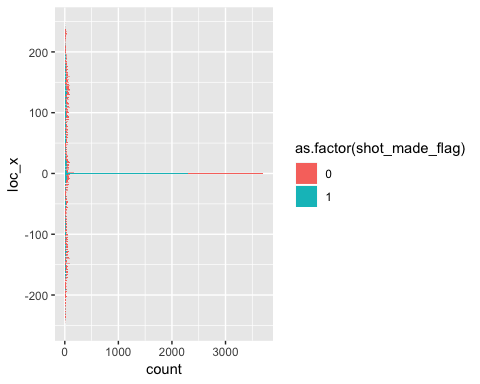


## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

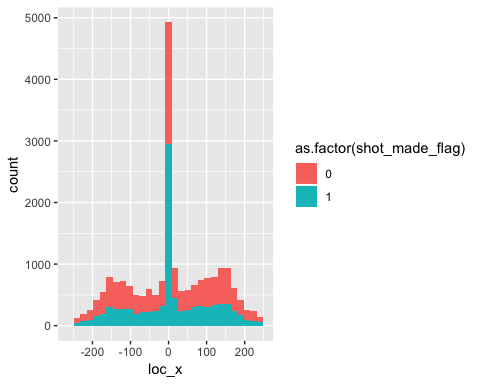


## <ggproto object: Class CoordFlip, CoordCartesian, Coord, gg>  
## aspect: function  
## backtransform\_range: function  
## clip: on  
## default: FALSE  
## distance: function  
## expand: TRUE  
## is\_free: function  
## is\_linear: function  
## labels: function  
## limits: list  
## modify\_scales: function  
## range: function  
## render\_axis\_h: function  
## render\_axis\_v: function  
## render\_bg: function  
## render\_fg: function  
## setup\_data: function  
## setup\_layout: function  
## setup\_panel\_params: function  
## setup\_params: function  
## transform: function  
## super: <ggproto object: Class CoordFlip, CoordCartesian, Coord, gg>

### loc\_x

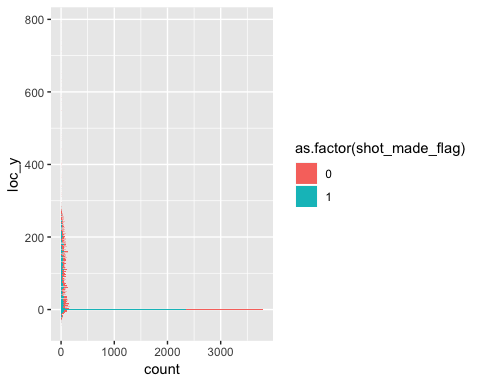


## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

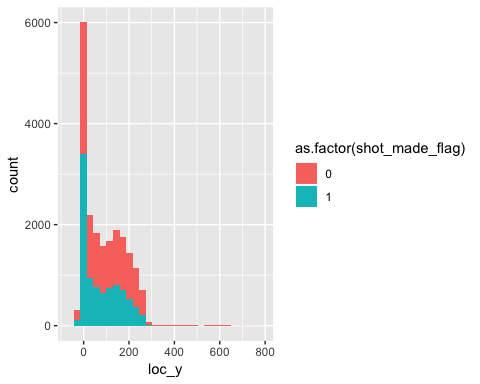


## <ggproto object: Class CoordFlip, CoordCartesian, Coord, gg>  
## aspect: function  
## backtransform\_range: function  
## clip: on  
## default: FALSE  
## distance: function  
## expand: TRUE  
## is\_free: function  
## is\_linear: function  
## labels: function  
## limits: list  
## modify\_scales: function  
## range: function  
## render\_axis\_h: function  
## render\_axis\_v: function  
## render\_bg: function  
## render\_fg: function  
## setup\_data: function  
## setup\_layout: function  
## setup\_panel\_params: function  
## setup\_params: function  
## transform: function  
## super: <ggproto object: Class CoordFlip, CoordCartesian, Coord, gg>

### loc\_y

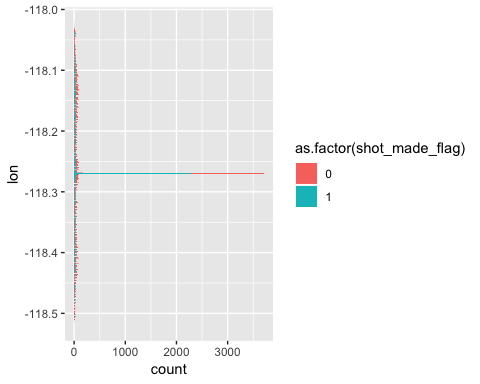


## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

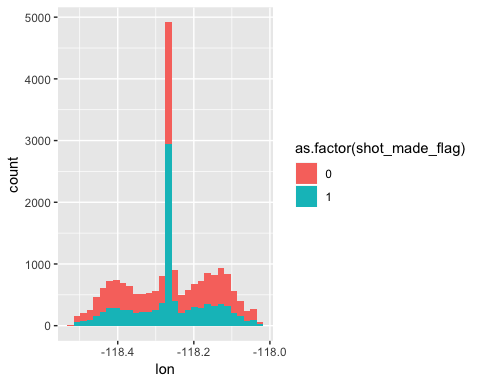


## <ggproto object: Class CoordFlip, CoordCartesian, Coord, gg>  
## aspect: function  
## backtransform\_range: function  
## clip: on  
## default: FALSE  
## distance: function  
## expand: TRUE  
## is\_free: function  
## is\_linear: function  
## labels: function  
## limits: list  
## modify\_scales: function  
## range: function  
## render\_axis\_h: function  
## render\_axis\_v: function  
## render\_bg: function  
## render\_fg: function  
## setup\_data: function  
## setup\_layout: function  
## setup\_panel\_params: function  
## setup\_params: function  
## transform: function  
## super: <ggproto object: Class CoordFlip, CoordCartesian, Coord, gg>

### lon

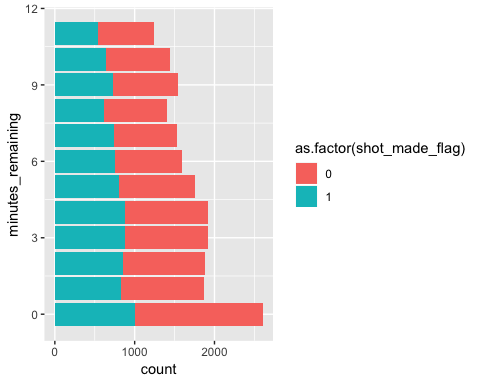


## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

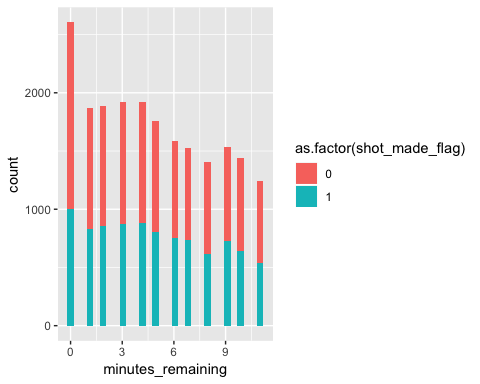


## <ggproto object: Class CoordFlip, CoordCartesian, Coord, gg>  
## aspect: function  
## backtransform\_range: function  
## clip: on  
## default: FALSE  
## distance: function  
## expand: TRUE  
## is\_free: function  
## is\_linear: function  
## labels: function  
## limits: list  
## modify\_scales: function  
## range: function  
## render\_axis\_h: function  
## render\_axis\_v: function  
## render\_bg: function  
## render\_fg: function  
## setup\_data: function  
## setup\_layout: function  
## setup\_panel\_params: function  
## setup\_params: function  
## transform: function  
## super: <ggproto object: Class CoordFlip, CoordCartesian, Coord, gg>

### minutes\_remaining

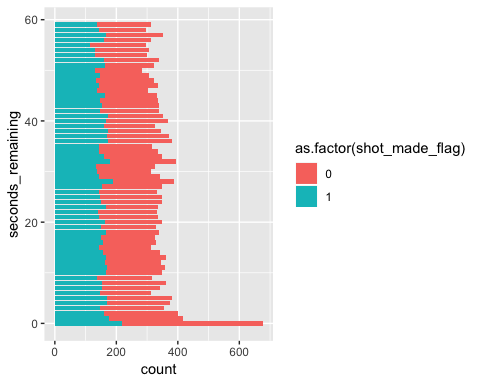


## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

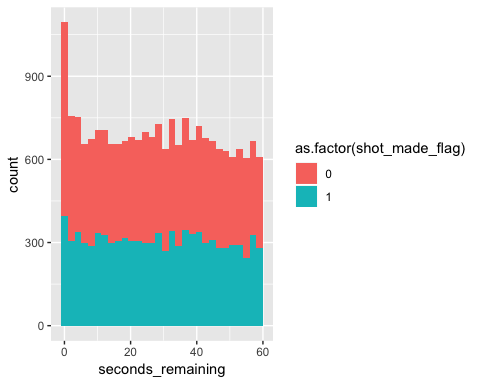


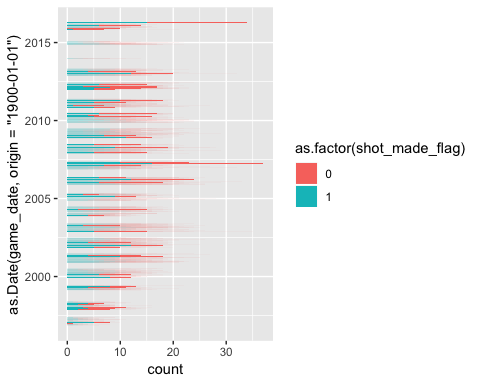
## <ggproto object: Class CoordFlip, CoordCartesian, Coord, gg>  
## aspect: function  
## backtransform\_range: function  
## clip: on  
## default: FALSE  
## distance: function  
## expand: TRUE  
## is\_free: function  
## is\_linear: function  
## labels: function  
## limits: list  
## modify\_scales: function  
## range: function  
## render\_axis\_h: function  
## render\_axis\_v: function  
## render\_bg: function  
## render\_fg: function  
## setup\_data: function  
## setup\_layout: function  
## setup\_panel\_params: function  
## setup\_params: function  
## transform: function  
## super: <ggproto object: Class CoordFlip, CoordCartesian, Coord, gg>

### seconds\_remaining

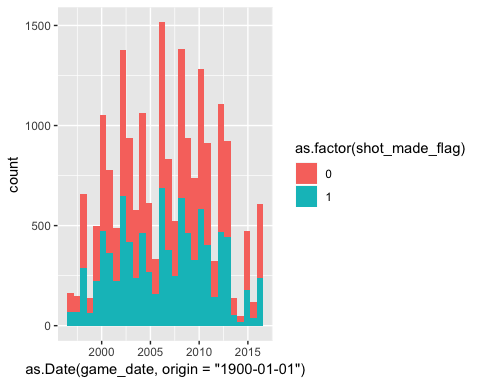


## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

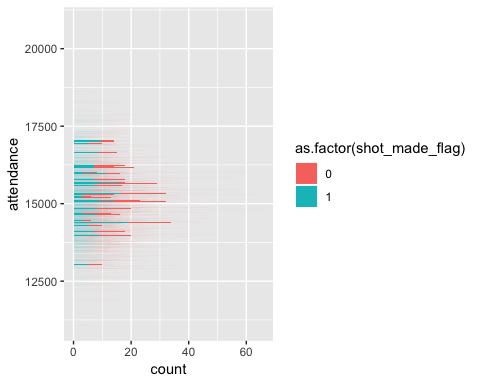
 ### game\_date



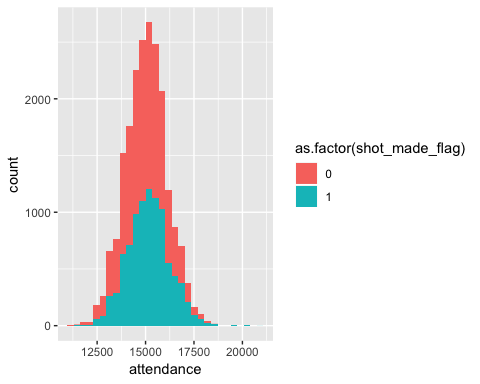
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



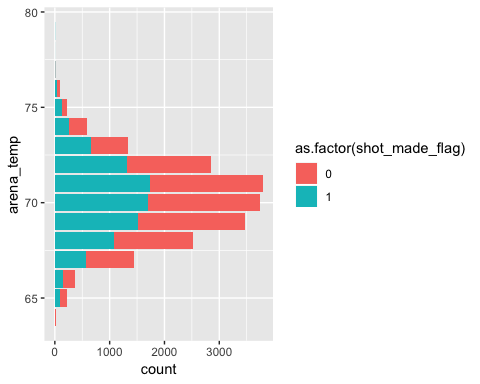
### attendance



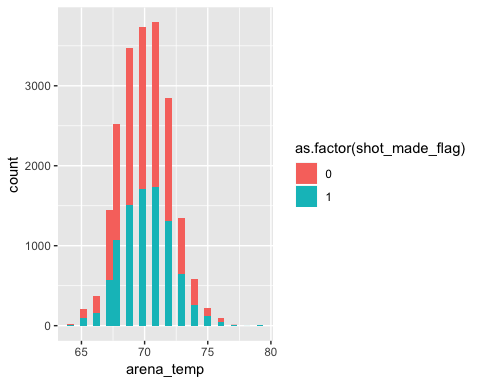
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



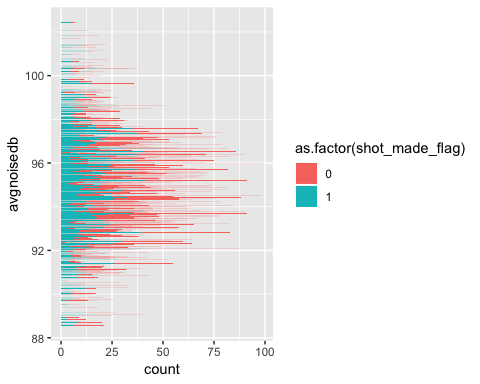
### arena\_temp



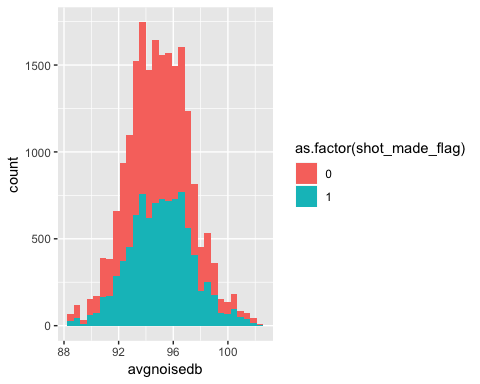
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



### avgnoisedb



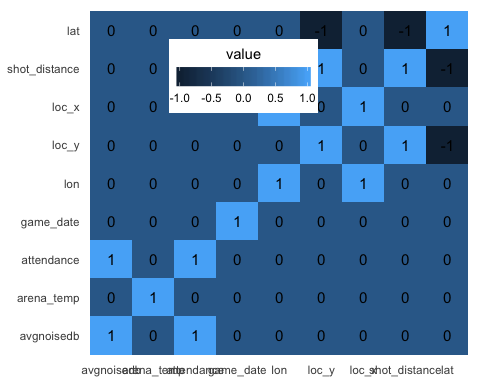
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



### Multicolineariy detection using VIF

## avgnoisedb arena\_temp attendance game\_date lon loc\_y loc\_x  
## avgnoisedb 1 0 1 0 0 0 0  
## arena\_temp 0 1 0 0 0 0 0  
## attendance 1 0 1 0 0 0 0  
## game\_date 0 0 0 1 0 0 0  
## lon 0 0 0 0 1 0 1  
## loc\_y 0 0 0 0 0 1 0  
## shot\_distance lat  
## avgnoisedb 0 0  
## arena\_temp 0 0  
## attendance 0 0  
## game\_date 0 0  
## lon 0 0  
## loc\_y 1 -1

## Var1 Var2 value  
## 1 avgnoisedb avgnoisedb 1  
## 2 arena\_temp avgnoisedb 0  
## 3 attendance avgnoisedb 1  
## 4 game\_date avgnoisedb 0  
## 5 lon avgnoisedb 0  
## 6 loc\_y avgnoisedb 0

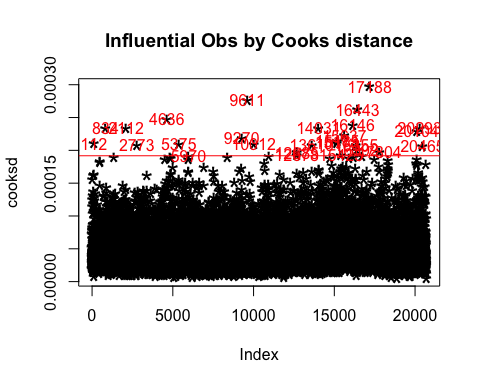


## arena\_temp game\_date   
## 81802.84 204061.55   
## attendance arena\_temp:game\_date   
## 498164.70 290160.27   
## arena\_temp:attendance game\_date:attendance   
## 609532.82 670177.98   
## arena\_temp:game\_date:attendance   
## 788182.21

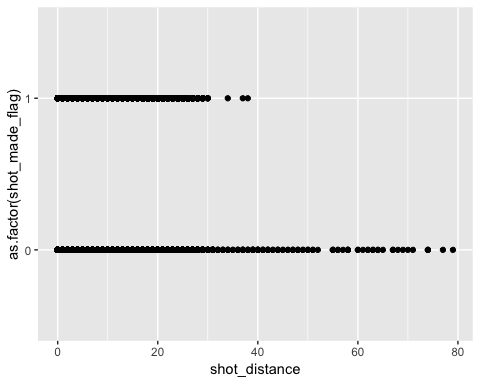
## Model :  
## shot\_made\_flag ~ avgnoisedb + arena\_temp + attendance + game\_date +   
## seconds\_remaining + minutes\_remaining + lon + lat + game\_event\_id

## avgnoisedb arena\_temp attendance game\_date   
## 1.350766 1.005825 1.360433 1.014508   
## seconds\_remaining minutes\_remaining lon lat   
## 1.002014 1.080305 1.001722 1.016855   
## game\_event\_id   
## 1.081280

### Outlier detection using cook’s D value



### Relation of shot made with distance and playoffs



## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

