Lecture Notes 2/1/2023

Prof. Bisbee, Vanderbilt University

2023-02-01

Load the data

```
require(tidyverse)
```

```
## Loading required package: tidyverse
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──
## ✓ ggplot2 3.3.6     ✓ purrr   0.3.4
## ✓ tibble  3.1.7     ✓ dplyr   1.0.9
## ✓ tidyr   1.2.0     ✓ stringr 1.4.0
## ✓ readr   2.1.2     ✓ forcats 0.5.1
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✕ dplyr::filter() masks stats::filter()
## ✕ dplyr::lag()    masks stats::lag()
```

```
nba <- read_rds('https://github.com/jbisbee1/DS1000_S2023/blob/main/Lectures/4_Uni_Multivariate/data/nba_pl
ayers_2018.Rds?raw=true')

nba
```

```
## # A tibble: 530 × 37
##    namePlayer      idPlayer slugSeason numberPlayerSea… isRookie slugTeam idTeam
##    <chr>              <dbl> <chr>                 <dbl> <lgl>    <chr>     <dbl>
##  1 LaMarcus Aldri…   200746 2018-19                  12 FALSE    SAS      1.61e9
##  2 Quincy Acy        203112 2018-19                   6 FALSE    PHX      1.61e9
##  3 Steven Adams      203500 2018-19                   5 FALSE    OKC      1.61e9
##  4 Alex Abrines      203518 2018-19                   2 FALSE    OKC      1.61e9
##  5 Bam Adebayo      1628389 2018-19                   1 FALSE    MIA      1.61e9
##  6 Rawle Alkins     1628959 2018-19                   0 TRUE     CHI      1.61e9
##  7 Grayson Allen    1628960 2018-19                   0 TRUE     UTA      1.61e9
##  8 Deng Adel        1629061 2018-19                   0 TRUE     CLE      1.61e9
##  9 Jaylen Adams     1629121 2018-19                   0 TRUE     ATL      1.61e9
## 10 DeVaughn Akoon…  1629152 2018-19                   0 TRUE     DEN      1.61e9
## # … with 520 more rows, and 30 more variables: gp <dbl>, gs <dbl>, fgm <dbl>,
## #   fga <dbl>, pctFG <dbl>, fg3m <dbl>, fg3a <dbl>, pctFG3 <dbl>, pctFT <dbl>,
## #   fg2m <dbl>, fg2a <dbl>, pctFG2 <dbl>, agePlayer <dbl>, minutes <dbl>,
## #   ftm <dbl>, fta <dbl>, oreb <dbl>, dreb <dbl>, treb <dbl>, ast <dbl>,
## #   stl <dbl>, blk <dbl>, tov <dbl>, pf <dbl>, pts <dbl>, urlNBAAPI <chr>,
## #   n <int>, org <fct>, country <chr>, idConference <int>
```

`summary()` function

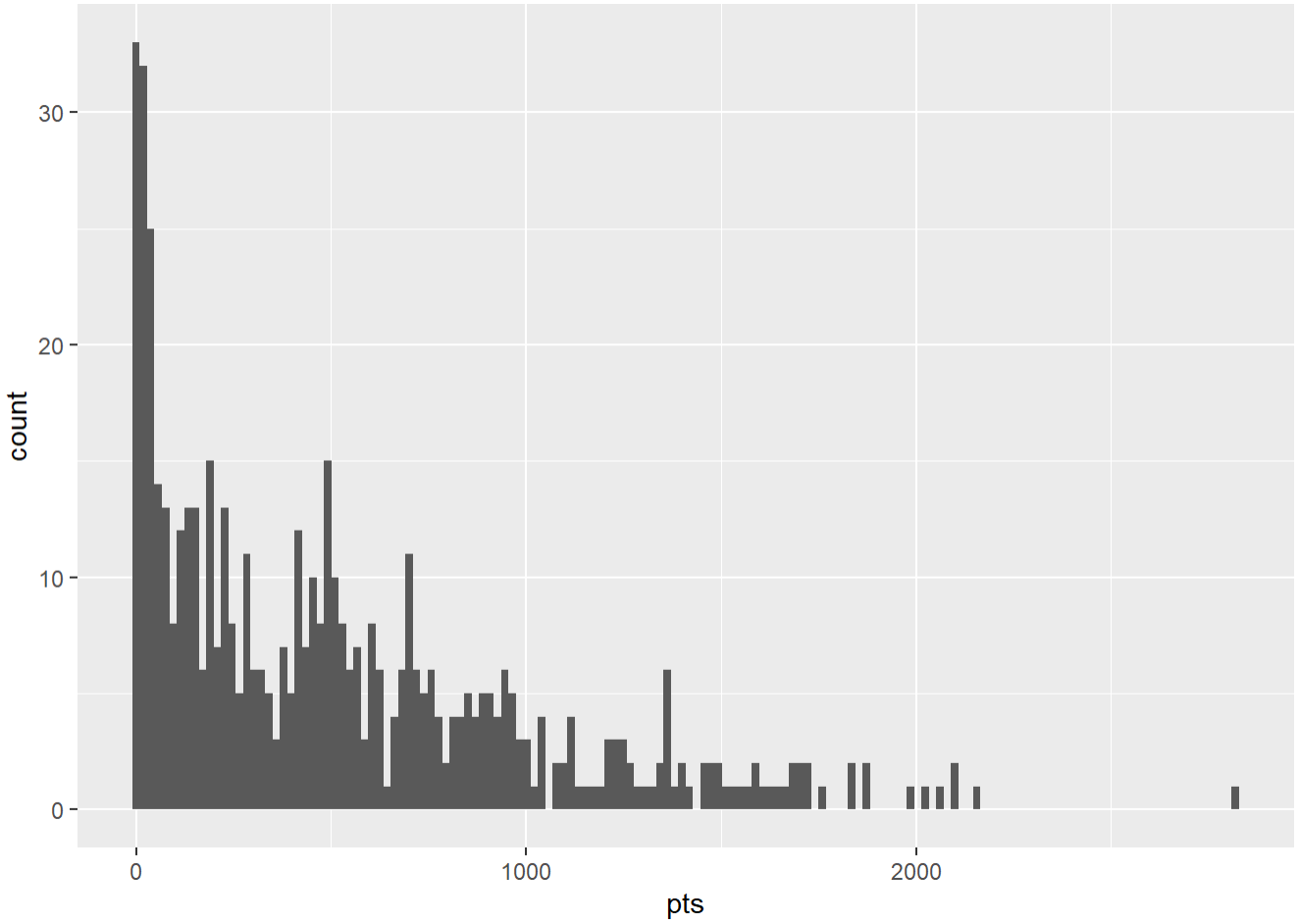
```
# nba %>% select(pts)
summary(nba$pts)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.0   115.0   419.0   516.2   759.5  2818.0
```

Univariate Visualization

- Using `geom_histogram()`

```
nba %>%
  ggplot(aes(x = pts)) +
  geom_histogram(bins = 150)
```

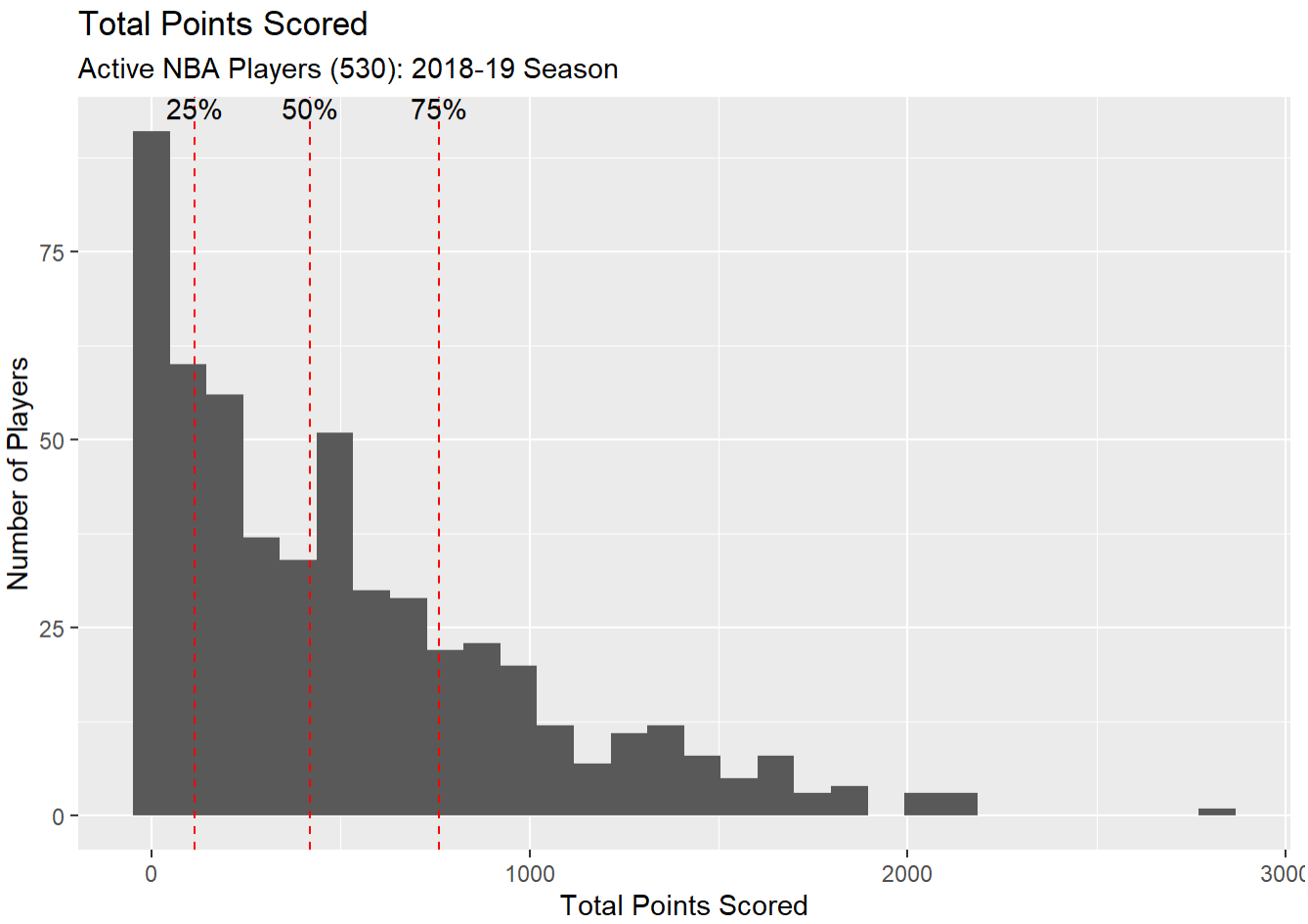


Adding

summary stats visually

```
nba %>%
  ggplot(aes(x = pts)) +
  geom_histogram() +
  geom_vline(xintercept = quantile(x = nba$pts,c(.25,.5,.75)),
             color = 'red',linetype = 'dashed') +
  annotate(geom = 'text',x = quantile(x = nba$pts,c(.25,.5,.75)),
           y = Inf,label = c('25%','50%','75%'),vjust = 1) +
  labs(title = 'Total Points Scored',
       subtitle = "Active NBA Players (530): 2018-19 Season",
       x = "Total Points Scored",
       y = "Number of Players")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



The Process

Step 1: Look

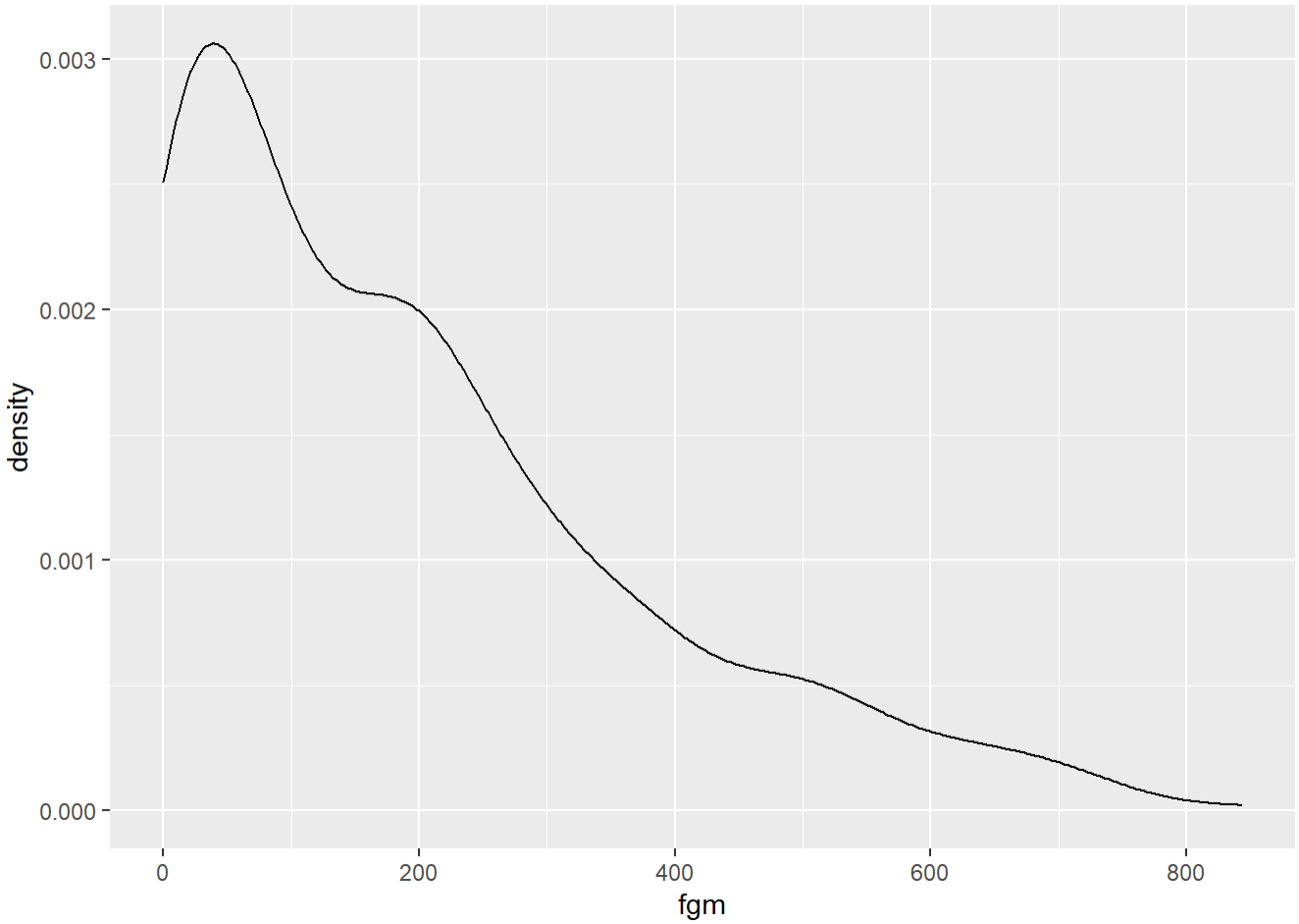
```
nba %>%
  select(fgm)
```

```
## # A tibble: 530 × 1
##      fgm
##    <dbl>
##  1   684
##  2     4
##  3   481
##  4    56
##  5   280
##  6    13
##  7    67
##  8    11
##  9    38
## 10     3
## # … with 520 more rows
```

```
summary(nba$fgm)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00   43.25  157.00  190.68  283.50  843.00
```

```
nba %>%
  ggplot(aes(x = fgm)) +
  geom_density()
```



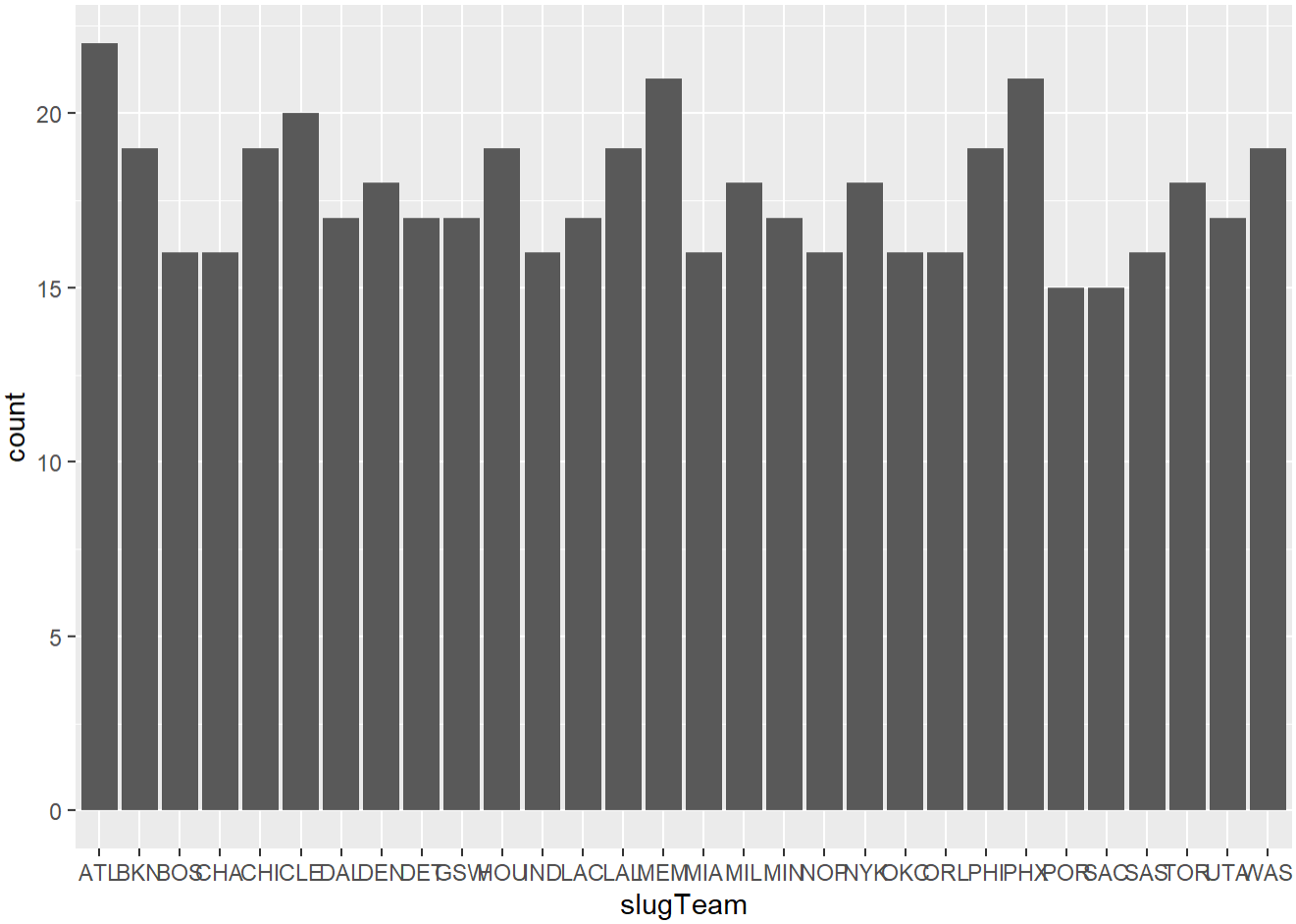
```
  # geom_histogram(bins = 100)

nba %>%
  select(slugTeam)
```

```
## # A tibble: 530 × 1
##    slugTeam
##    <chr>
##  1 SAS
##  2 PHX
##  3 OKC
##  4 OKC
##  5 MIA
##  6 CHI
##  7 UTA
##  8 CLE
##  9 ATL
## 10 DEN
## # … with 520 more rows
```

```
summary(nba$slugTeam)
```

```
##    Length     Class      Mode
##       530 character character
```

```
nba %>%
  ggplot(aes(x = slugTeam)) +
  geom_bar()
```


count
20
15
10
5
0
ATL BKN BOS CHA CHI CLE DAL DEN DET GSW HOU IND LAC LAL MEM MIA MIL MIN NOP NYK OKC ORL PHI PHX POR SAC SAS TOR UTA WAS
slugTeam