

# NCAA Basketball Coaches Salary Data 2013

Sumedh Deepak Patkar

2023-02-04

## Summary

The National Collegiate Athletic Association (NCAA) published in 2013 the Salary data of Men's basketball team coaches. The excel sheet can be found on this [link](https://github.com/Sumedh-Patkar/ST516/blob/main/data/1_2013_NCAA_basketball_coaches_salaries.xls?raw=true) (https://github.com/Sumedh-Patkar/ST516/blob/main/data/1\_2013\_NCAA\_basketball\_coaches\_salaries.xls?raw=true).

The sheet contains salaries paid to the coaches of 62 teams in a recent NCAA men's Division 1 basketball tournament. We consider these 62 salaries as *sample data* from the population of 346 NCAA Division 1 Men's basketball coaches in the year the tournament was held.

In this RMarkdown sheet, we will load, visualize, summarize the data. We will then calculate a **98% confidence interval** for the coach salaries of the population.

## Loading the data

```
library("dplyr")
library("readxl")
library("ggplot2")
library("scales")

# Set the working directory to code directory
# You may need to change this variable if you try running this code on your machine
current_working_dir = "~/Development/ST516/1.NC_AA_Basketball_Coaches_Salaries"
setwd(current_working_dir)

# Read the excel file using readxl package function 'read_excel()'
salaries_xl <- readxl::read_excel("../data/1_2013_NCAA_basketball_coaches_salaries.xls", sheet = 1)

salaries_xl <- salaries_xl %>%
  rename(athletic_conference = CONF, school=SCHOOL, head_coach = `HEAD COACH`, school_pay = `SCHOOL PAY`,
         other_pay = `OTHER PAY`, total_pay = `TOTAL PAY`, max_bonus = `MAX BONUS`)
```

## Visualizing the data

Tabular view (First 6 rows only)

```
#Tabular representation of data
knitr::kable(head(salaries_xl))
```

RANK	school	athletic_conference	head_coach	school_pay	other_pay	total_pay	max_bonus
1	Duke	ACC	Mike Krzyzewski	7233976	0	7233976	0
2	Louisville	Big East	Rick Pitino	4078327	895016	4973343	725000
3	Kansas	Big 12	Bill Self	4750763	210000	4960763	525000
4	Michigan State	Big Ten	Tom Izzo	3220769	525000	3745769	350000
5	Florida	SEC	Billy Donovan	3689200	0	3689200	471500
6	Ohio State	Big Ten	Thad Matta	3087000	107000	3194000	410000

## Summary of the data

```
# We skip the categorical attributes here. We will summarize some of them graphically below.
salaries_xl %>%
  dplyr::select(-'RANK', -'head_coach', -'school', -'athletic_conference') %>%
  summary() %>%
  knitr::kable()
```

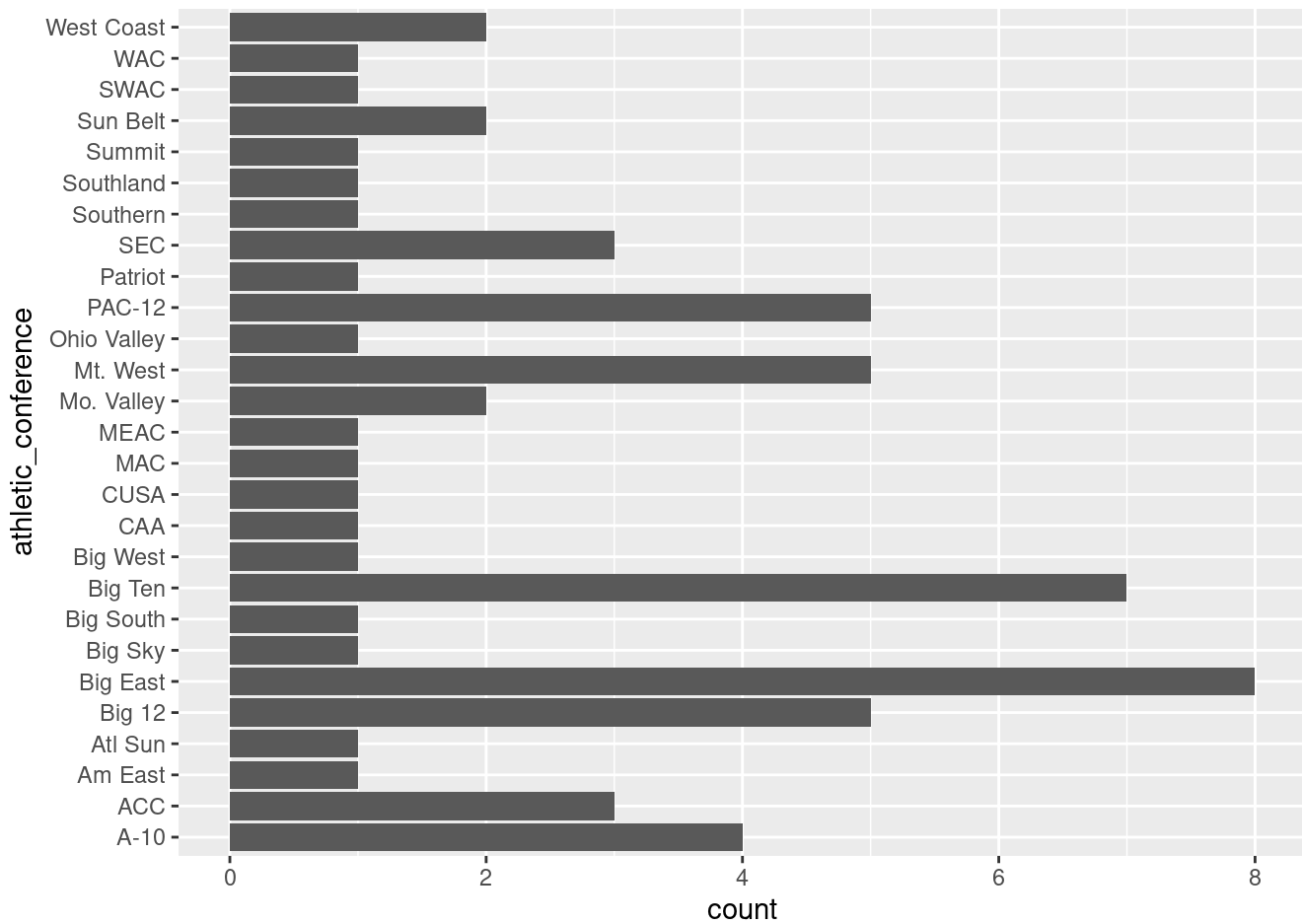
school_pay	other_pay	total_pay	max_bonus
Min. : 115000	Min. : 0	Min. : 115000	Min. : 0
1st Qu.: 425308	1st Qu.: 0	1st Qu.: 428140	1st Qu.: 0
Median :1214600	Median : 0	Median :1233850	Median : 201350
Mean :1421056	Mean : 44696	Mean :1465752	Mean : 367311
3rd Qu.:1938894	3rd Qu.: 10375	3rd Qu.:1938894	3rd Qu.: 621094
Max. :7233976	Max. :895016	Max. :7233976	Max. :2600000

## Graphical summary views

### 1. Representation of teams from various athletic conferences

Following chart shows the number of teams belonging to various athletic conferences ([https://en.wikipedia.org/wiki/List\\_of\\_NCAA\\_conferences](https://en.wikipedia.org/wiki/List_of_NCAA_conferences))

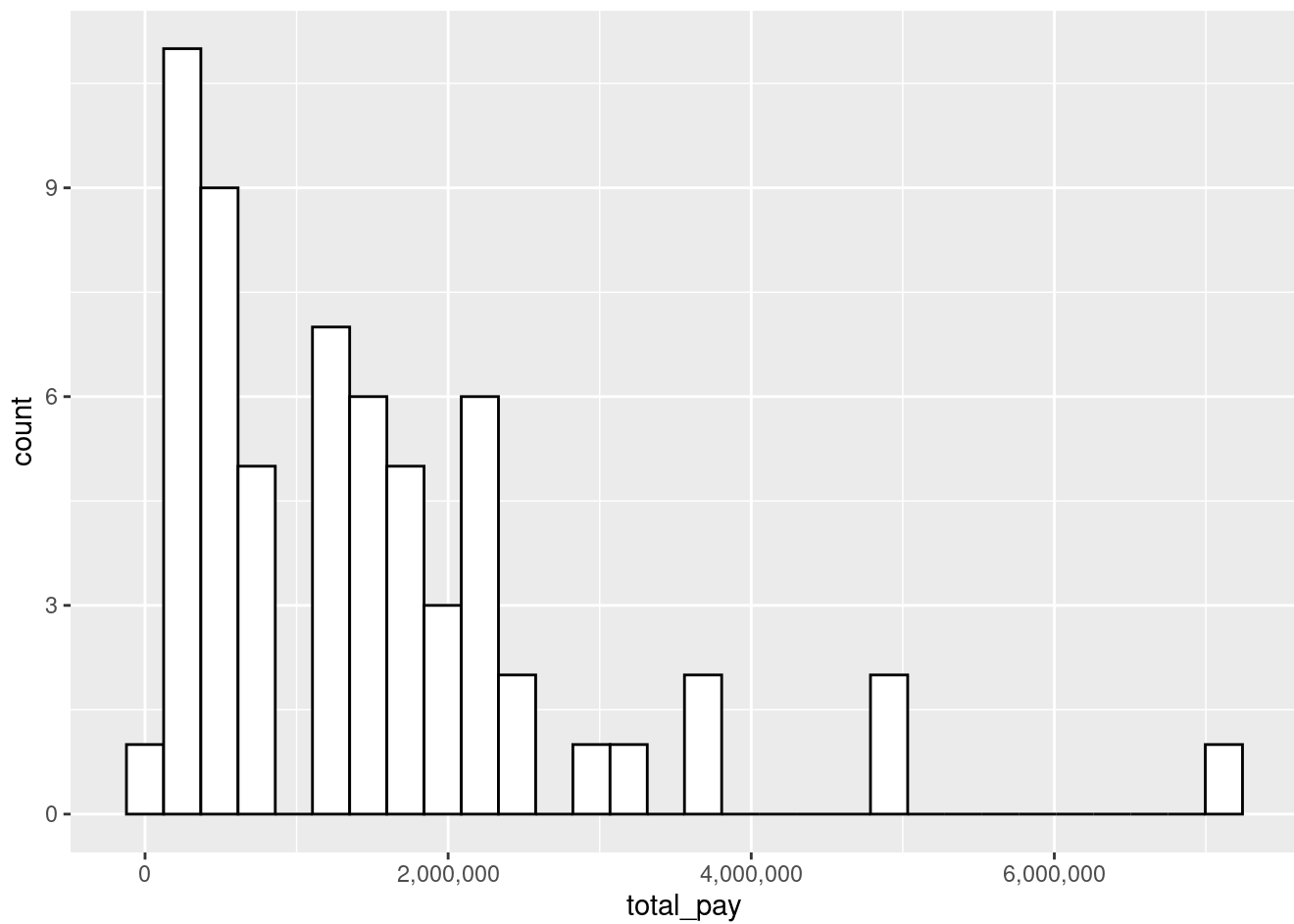
```
ggplot(data = salaries_xl) +
  geom_bar(mapping = aes(y = athletic_conference))
```



## 2. Histogram of Total Pay

Following chart shows the histogram of the total pay column

```
ggplot(data = salaries_xl) +  
  geom_histogram(mapping = aes(x = total_pay), colour="black", fill="white") +  
  scale_x_continuous(labels = comma)
```



Notice the outlier in the Total Pay.