R Notebook

```
# Name: Arjun Bhan
library("tidyverse")
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.6     v dplyr   1.0.4
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
FireFighters <- read.csv(file.choose(), stringsAsFactors = FALSE)
```

```
FireFighters
```

X <int>	Candidate <int>	Race <chr>	Position <chr>	Exam <chr>	score <dbl>
1	1	W	Captain	Oral	89.52
2	1	W	Captain	Written	95.00
3	2	W	Captain	Oral	80.00
4	2	W	Captain	Written	95.00
5	3	W	Captain	Oral	82.38
6	3	W	Captain	Written	87.00
7	4	W	Captain	Oral	88.57
8	4	W	Captain	Written	76.00
9	5	W	Captain	Oral	76.19
10	5	W	Captain	Written	84.00

1-10 of 236 rows Previous **1** 2 3 4 5 6 … 24 Next

 # The issue about this data is that the information in it is repeated in different rows. We need to make sure each row represent a separate unique data point to make the data set tidy. We can d o those by combing data that shares identical information. This can be done by placing the diffe rent exam type values as separate columns. By doing this we will reduce the redundancy of the da ta without losing important information. We can accomplish this task with the spread function to separate the different exam values into different columns.

```
FireFightersTidy<- FireFighters %>% select(-X) %>% spread(key = "Exam", value = "score")
FireFightersTidy
```

Candidate <int>	Race <chr>	Position <chr>	Oral <dbl>	Written <dbl>
1	W	Captain	89.52	95
2	W	Captain	80.00	95
3	W	Captain	82.38	87
4	W	Captain	88.57	76
5	W	Captain	76.19	84
6	H	Captain	76.19	82
7	W	Captain	76.19	82
8	H	Captain	70.00	84
9	W	Captain	73.81	81
10	W	Captain	84.29	72

1-10 of 118 rows Previous **1** 2 3 4 5 6 … 12 Next

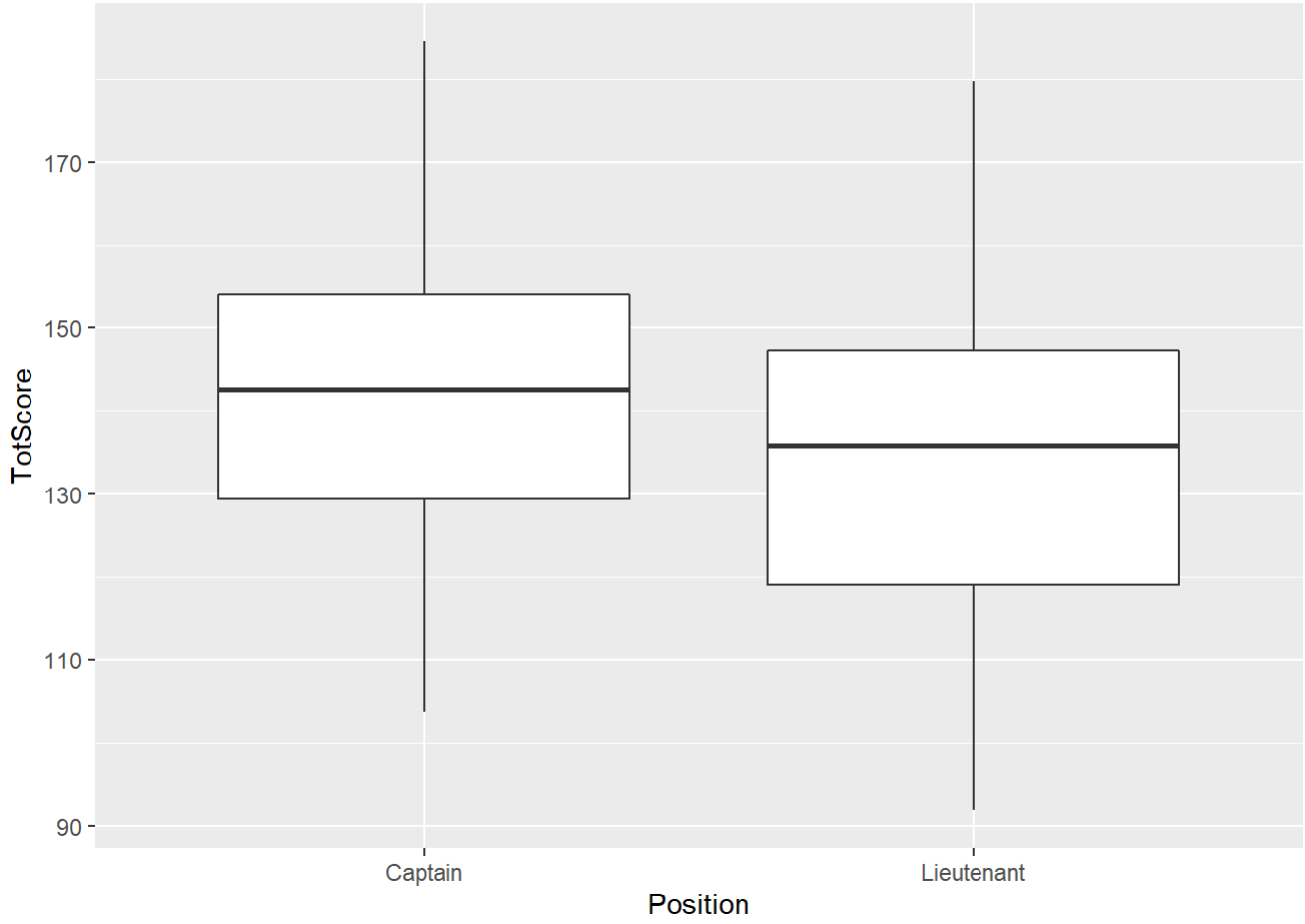
```
FireFightersTidy<- FireFightersTidy %>% mutate(TotScore = Oral+Written)
FireFightersTidy
```

Candidate <int>	Race <chr>	Position <chr>	Oral <dbl>	Written <dbl>	TotScore <dbl>
1	W	Captain	89.52	95	184.52
2	W	Captain	80.00	95	175.00
3	W	Captain	82.38	87	169.38
4	W	Captain	88.57	76	164.57
5	W	Captain	76.19	84	160.19
6	H	Captain	76.19	82	158.19
7	W	Captain	76.19	82	158.19
8	H	Captain	70.00	84	154.00

Candidate <int>	Race <chr>	Position <chr>	Oral <dbl>	Written <dbl>	TotScore <dbl>
9	W	Captain	73.81	81	154.81
10	W	Captain	84.29	72	156.29

1-10 of 118 rows Previous **1** 2 3 4 5 6 ... 12 Next

```
library(ggplot2)
ggplot(FireFightersTidy, aes(x=Position, y=TotScore))+geom_boxplot()
```




```
# This figure is showing that captains tend to have higher overall exam scores than lieutenants. This makes sense as captain are higher ranked than lieutenants. This indicates that higher scores help soldiers rank up from lieutenants to captains.
```