20BCE1025_Abhishek_N_N_Expeiment-10 Visualization using grammar of graphics

20BCE1025_Abhishek_N_N

2022-11-03

Use the newsurvey data obtained by cleaning 'na' values in survey data of MASS package and ggplot2 package to do the following:

```
library(MASS)
newSurvey=na.omit(survey)
head(newSurvey)
```

```
##      Sex Wr.Hnd NW.Hnd W.Hnd    Fold Pulse  Clap Exer Smoke Height      M.I
## 1 Female   18.5   18.0 Right  R on L    92  Left Some Never 173.00   Metric
## 2   Male   19.5   20.5  Left  R on L   104  Left None Regul 177.80 Imperial
## 5   Male   20.0   20.0 Right Neither    35 Right Some Never 165.00   Metric
## 6 Female   18.0   17.7 Right  L on R    64 Right Some Never 172.72 Imperial
## 7   Male   17.7   17.7 Right  L on R    83 Right Freq Never 182.88 Imperial
## 8 Female   17.0   17.3 Right  R on L    74 Right Freq Never 157.00   Metric
##      Age
## 1 18.250
## 2 17.583
## 5 23.667
## 6 21.000
## 7 18.833
## 8 35.833
```

1. Install the package ggplot2 and import it.

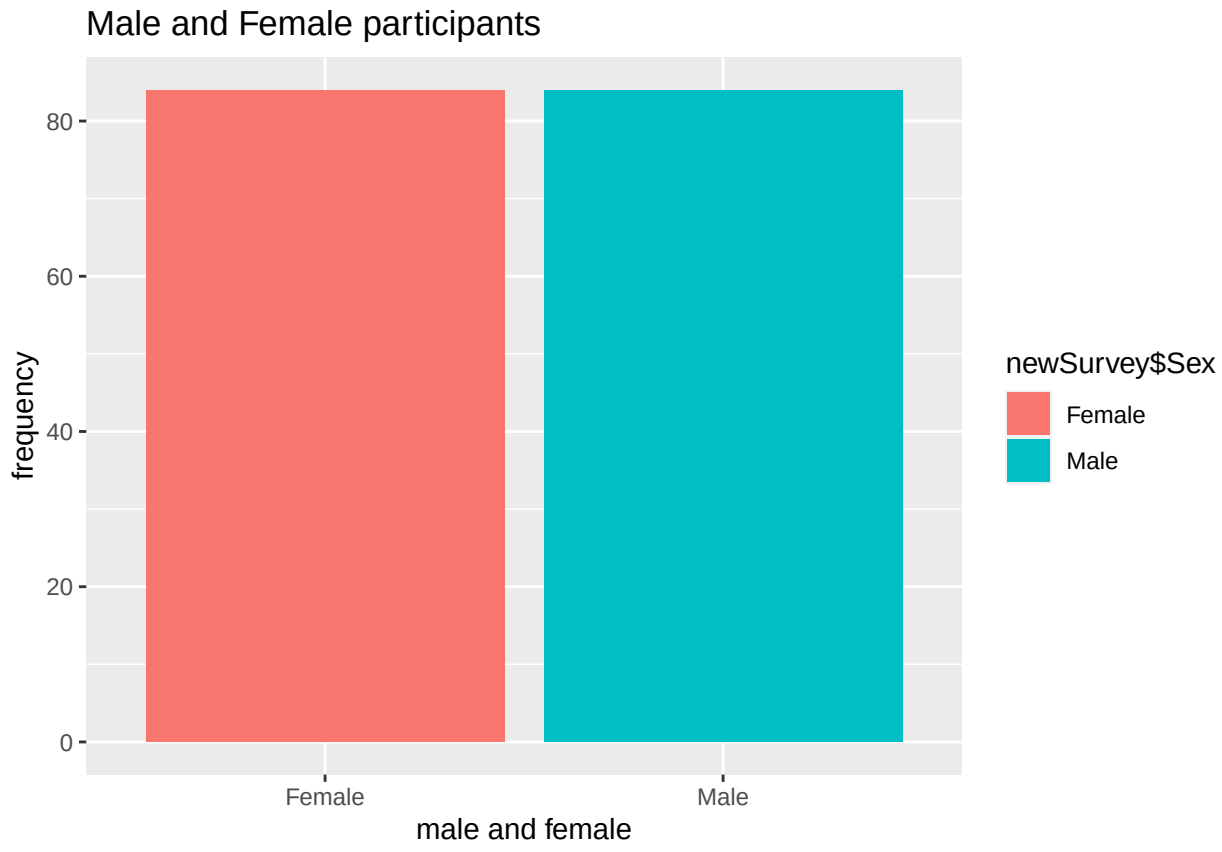
```
#install.packages("ggplot2")
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following object is masked from 'package:MASS':
##
##     select

## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

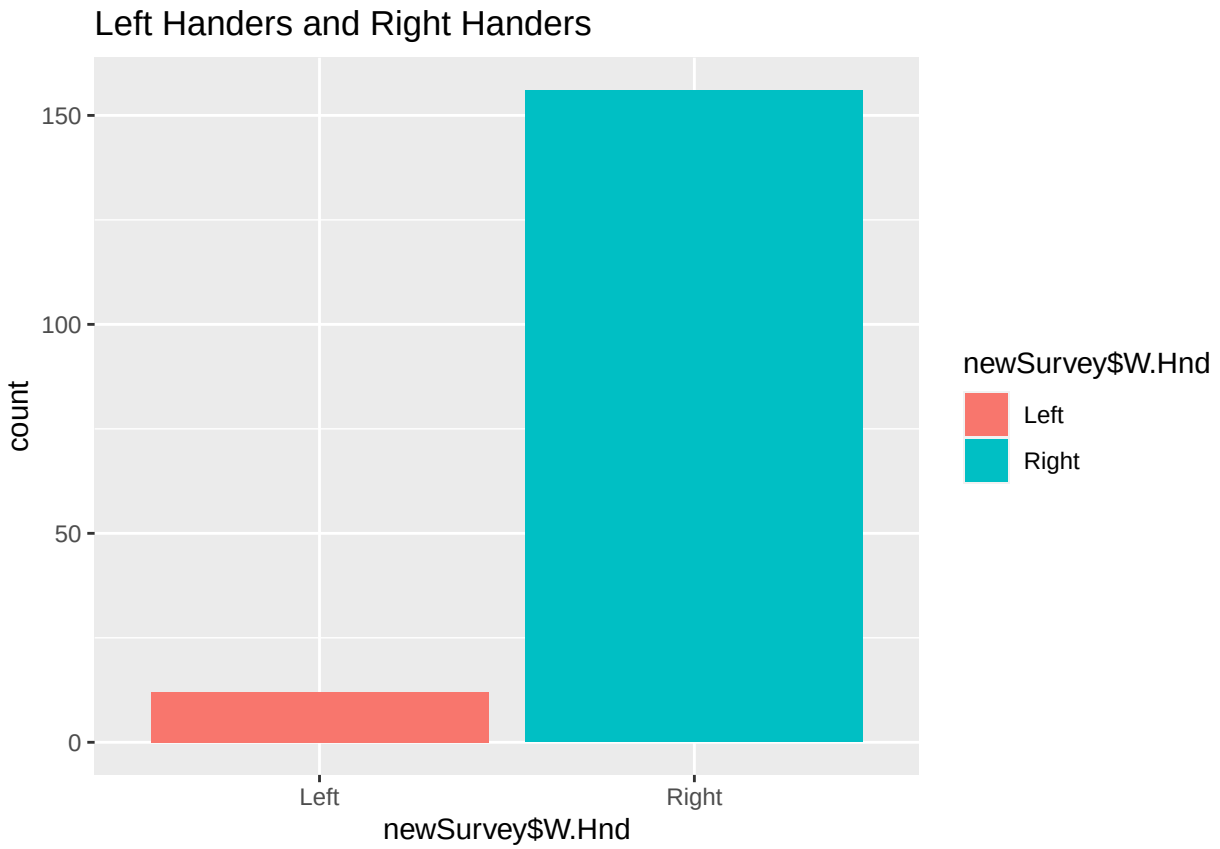
2. Plot a bar graph for the number of male and female participants in the survey. Provide the title as "Male and Female participants", y-axis label as "frequency" and specify the colours for the bars.

```
library(ggplot2)
ggplot(newSurvey)+
  geom_bar(aes(newSurvey$Sex,fill=newSurvey$Sex)) +
  ggtitle("Male and Female participants") +
  xlab("male and female") +
  ylab("frequency")
```



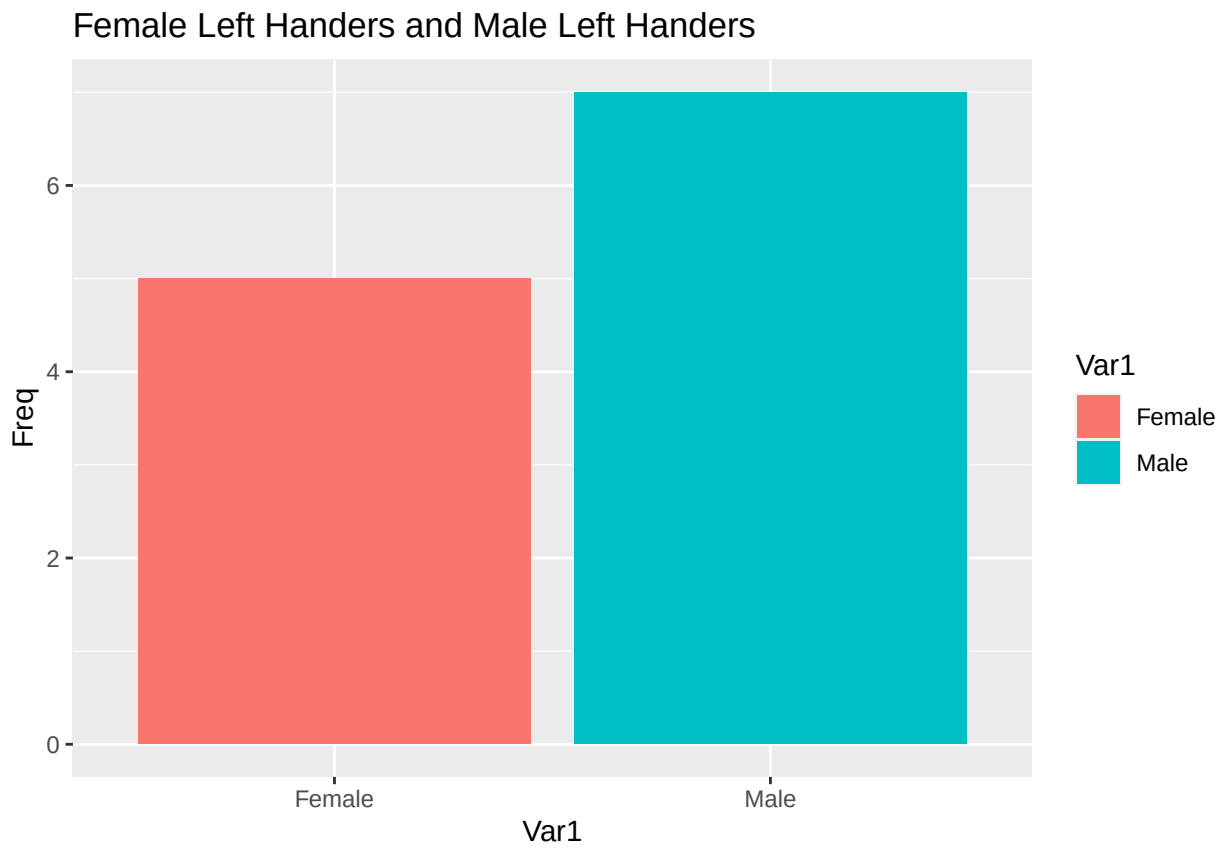
3. Plot a bar graph for the number of left handers and right handers in the survey. Provide the title as "Left Handers and Right Handers", y-axis label as "count" and specify the colours for the bars.

```
ggplot(newSurvey)+
  geom_bar(aes(newSurvey$W.Hnd,fill=newSurvey$W.Hnd)) +
  ggtitle("Left Handers and Right Handers") +
  ylab("count")
```



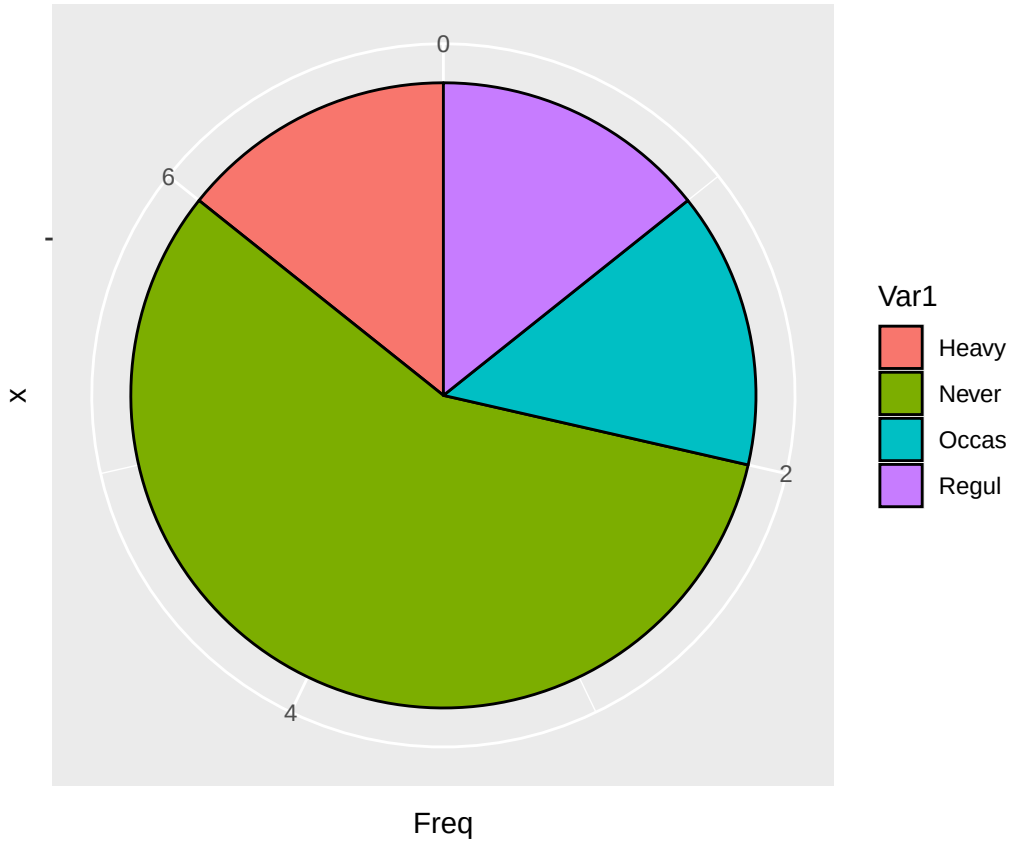
4. Plot the distribution between male left handers and female left handers using bar chart. Provide the title as "Female Left Handers and Male Left Handers , y-axis label as "count" and specify the colours for the bars.

```
male_left_handers= newSurvey[newSurvey$W.Hnd=="Left" & newSurvey$Sex=="Male",]
female_left_handers= newSurvey[newSurvey$W.Hnd=="Left" & newSurvey$Sex=="Female",]
custom_data = rbind(male_left_handers, female_left_handers)
ggplot(as.data.frame(table(custom_data$Sex)), aes(x=Var1, y = Freq,
fill=Var1)) +
geom_bar(stat="identity")+
labs(title="Female Left Handers and Male Left Handers")
```



5. Draw the distribution of smoking habits of male left handers using pie chart.

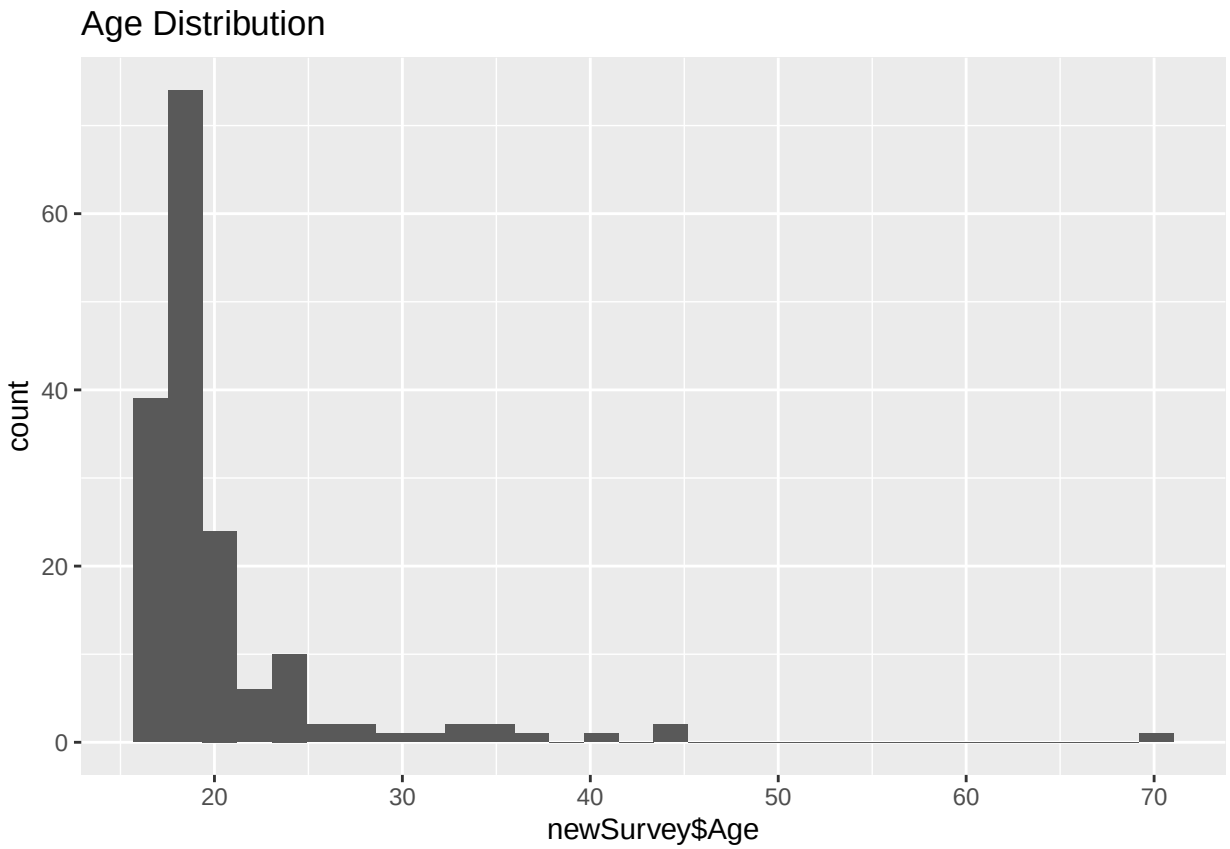
```
smoking_habits = ggplot(as.data.frame(table(male_left_handers$Smoke)),
aes(x = "", y =Freq, fill=Var1)) +
geom_bar(stat="identity")+
geom_col(color = "black") +
coord_polar("y", start=0)
smoking_habits
```



6. Draw the histogram of age distribution with the title as 'Age distribution' and xlabel as 'Age range' and ylabel as 'frequency'.

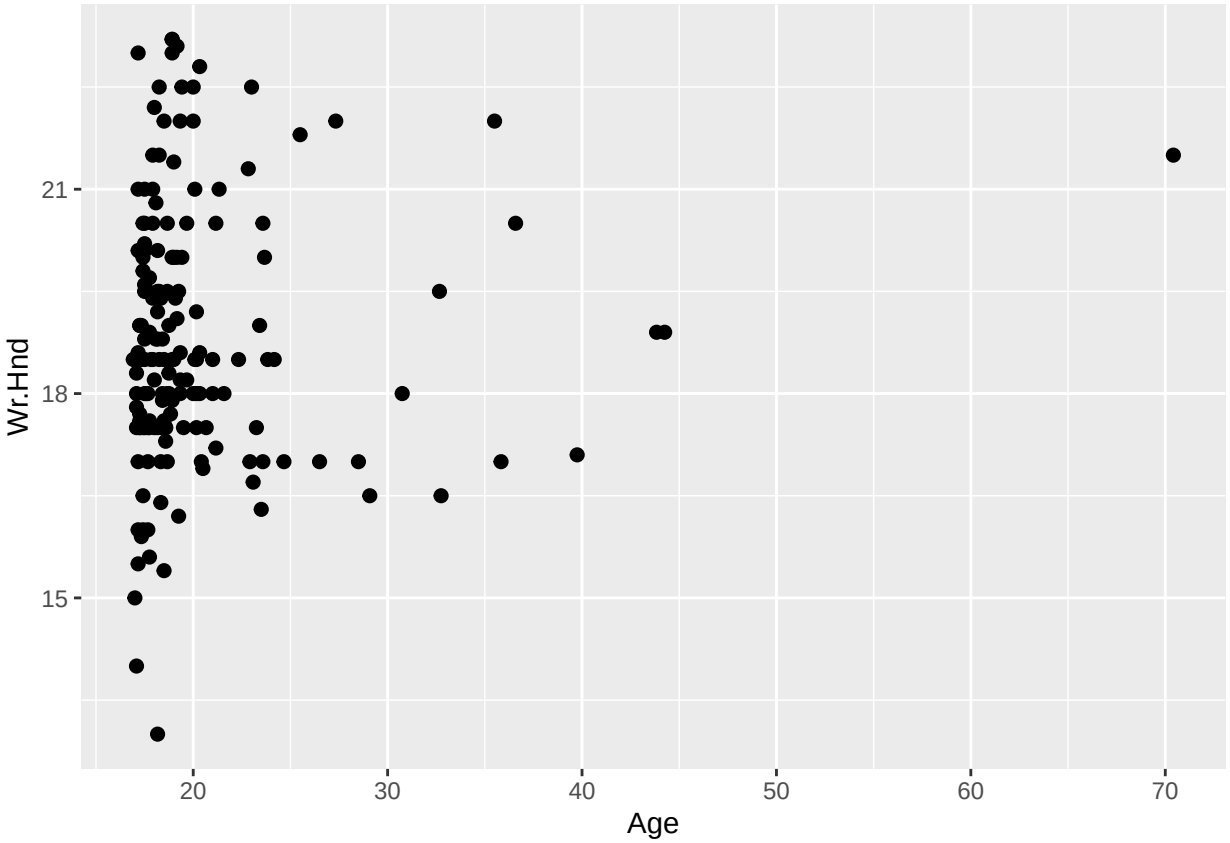
```
ggplot(newSurvey, aes(x=newSurvey$Age)) +
  geom_histogram() +
  stat_bin(bins = 30) +
  labs(title="Age Distribution", xlabel="Age Range", ylabel="Frequency")
```

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



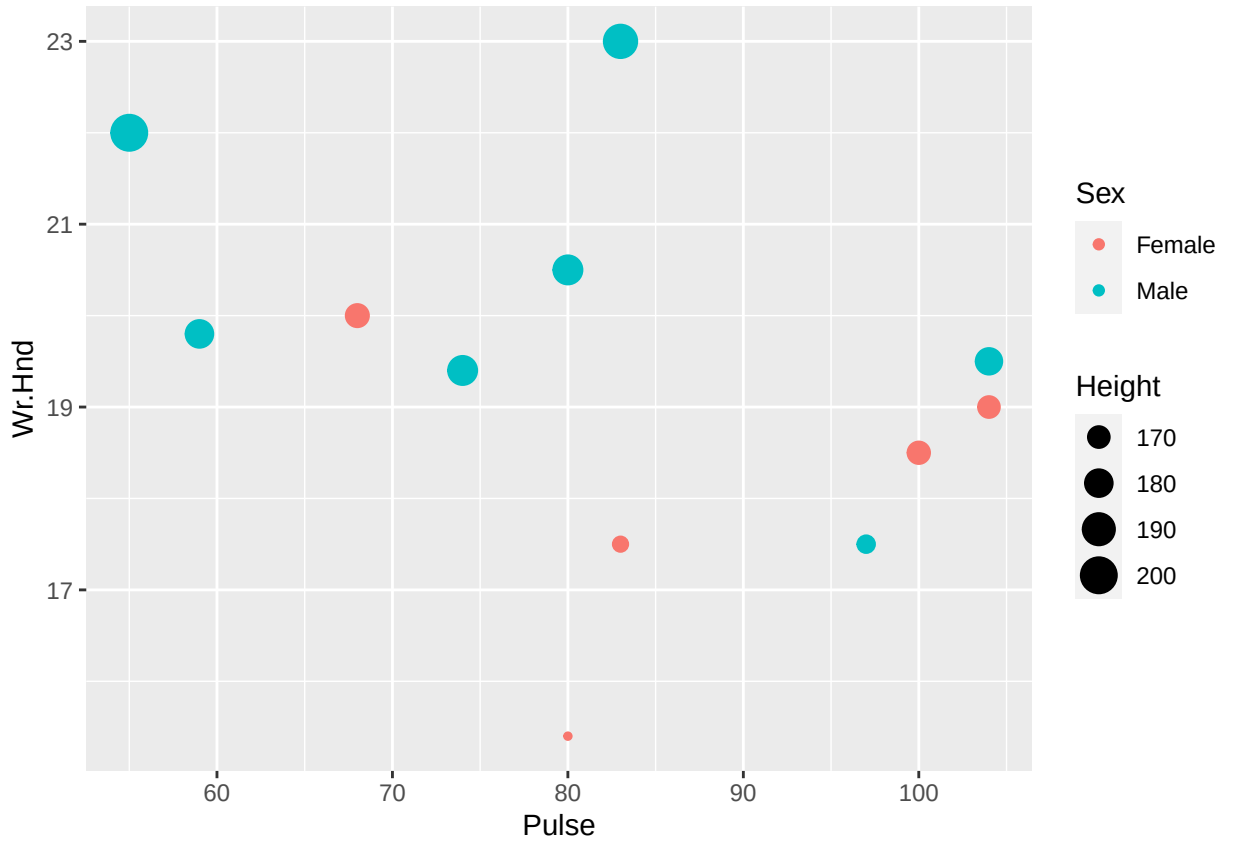
7. Reveal the relationship between the age and writing hand span using scatter plot.

```
ggplot(newSurvey, aes(x=Age, y=Wr.Hnd)) + geom_point(size=2)
```



8. Plot the distribution of writing hand span vs. pulse rate of left handers. Provide colour based on gender and vary the size of the point based on height of the student.

```
ggplot(custom_data, aes(x=Pulse, y=Wr.Hnd, color=Sex, size=Height)) + geom_point()
```



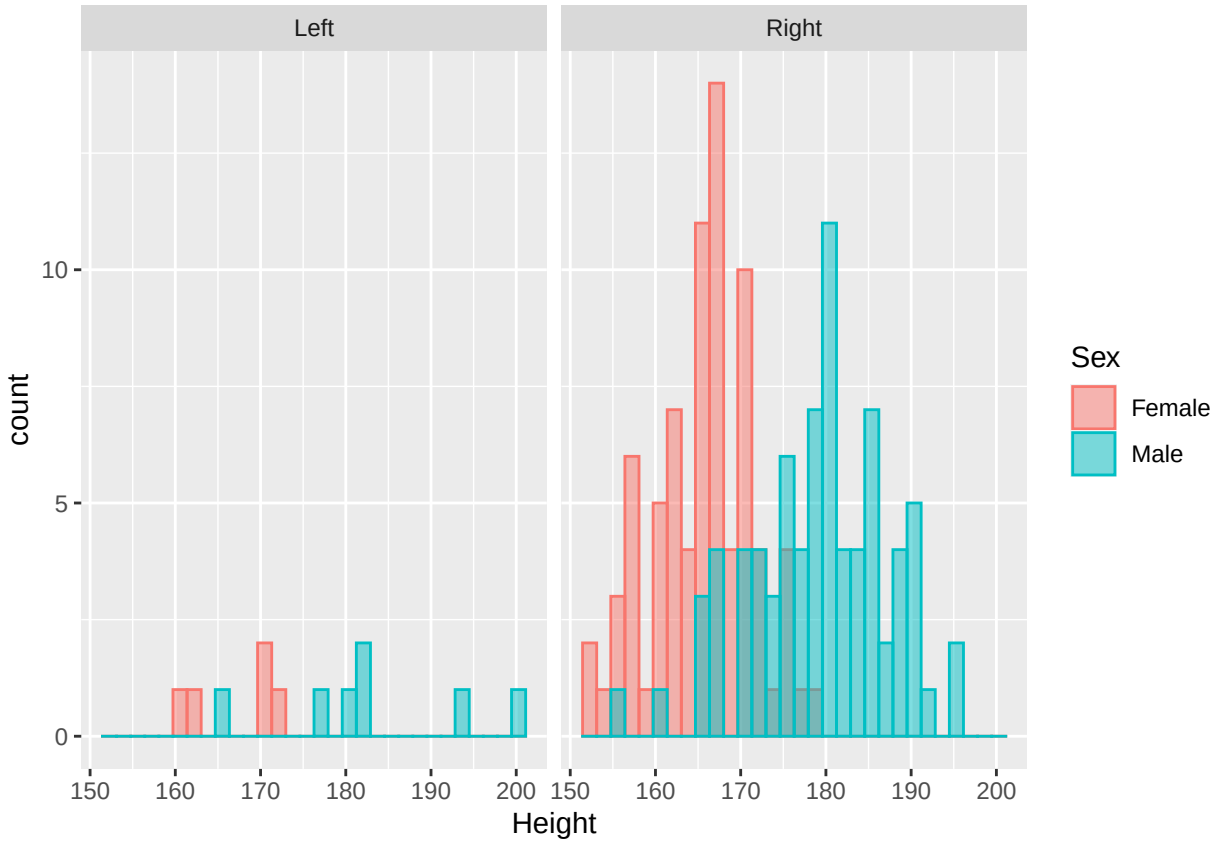
9. Plot the distribution of height of the students with filled colour based on gender with facet based on left and write handers.

```
head(newSurvey)
```

```
##      Sex Wr.Hnd NW.Hnd W.Hnd    Fold Pulse  Clap Exer Smoke Height      M.I
## 1 Female   18.5   18.0 Right  R on L    92  Left Some Never 173.00   Metric
## 2   Male   19.5   20.5  Left  R on L   104  Left None Regul 177.80 Imperial
## 5   Male   20.0   20.0 Right Neither    35 Right Some Never 165.00   Metric
## 6 Female   18.0   17.7 Right  L on R    64 Right Some Never 172.72 Imperial
## 7   Male   17.7   17.7 Right  L on R    83 Right Freq Never 182.88 Imperial
## 8 Female   17.0   17.3 Right  R on L    74 Right Freq Never 157.00   Metric
##      Age
## 1 18.250
## 2 17.583
## 5 23.667
## 6 21.000
## 7 18.833
## 8 35.833
```

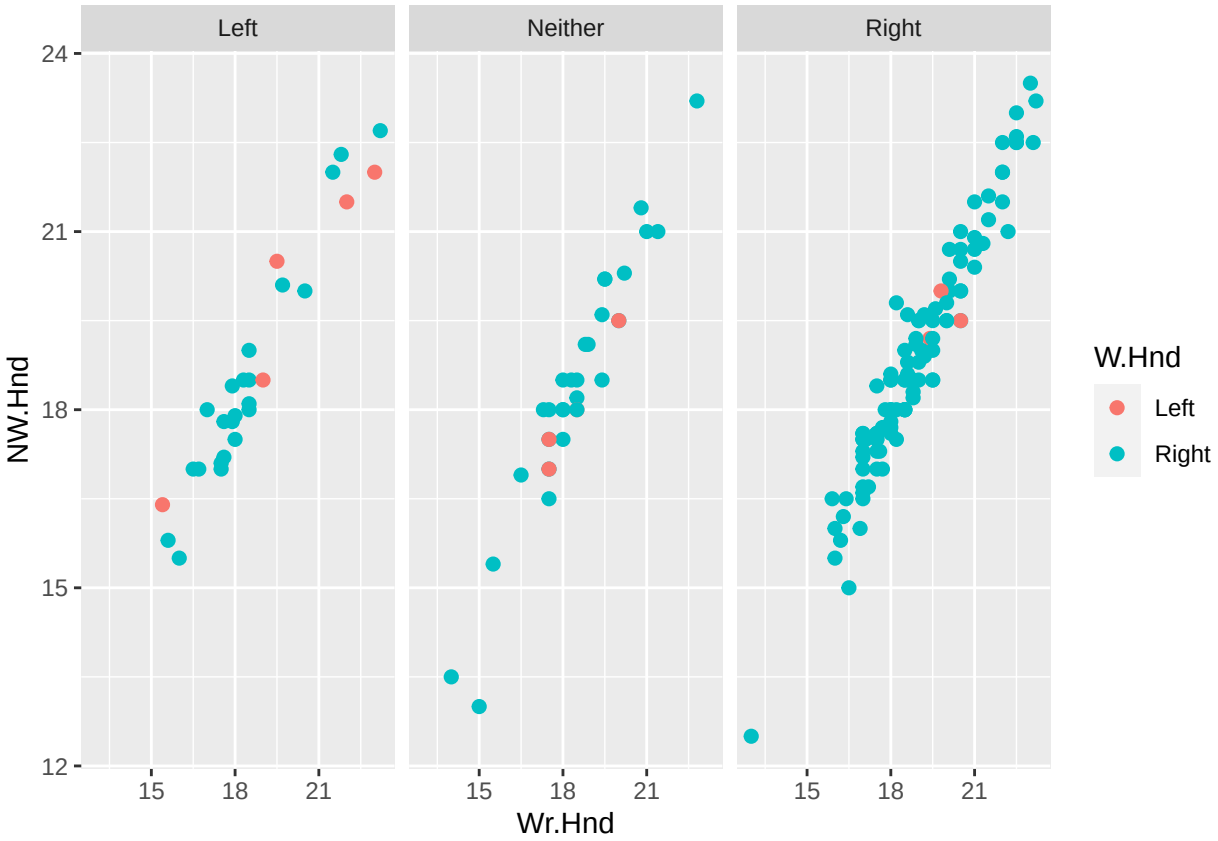
```
ggplot(newSurvey, aes(x=Height, color=Sex,fill=Sex)) +
  geom_histogram( alpha=0.5, position="identity") +
  facet_wrap(~W.Hnd)
```

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



10. Plot the trend of span of writing hand vs. non-writing hand coloured and grouped based on left and right handers with facet label based on clap.

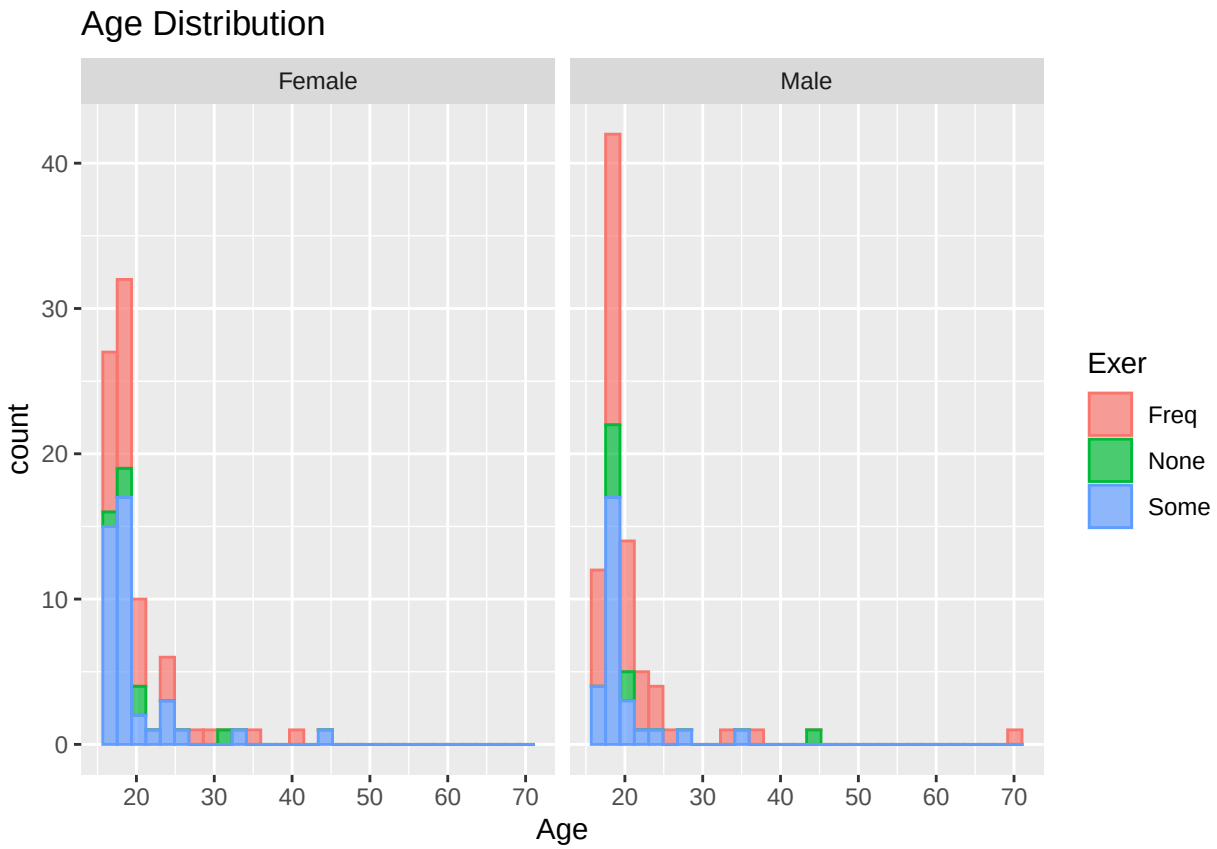
```
ggplot(newSurvey, aes(x=Wr.Hnd, y=NW.Hnd, color=W.Hnd, )) +
facet_grid(.~Clap) +
geom_point(size=2)
```



11. Plot the distribution of age of students based on categories under exercise with facet wrap based on gender.

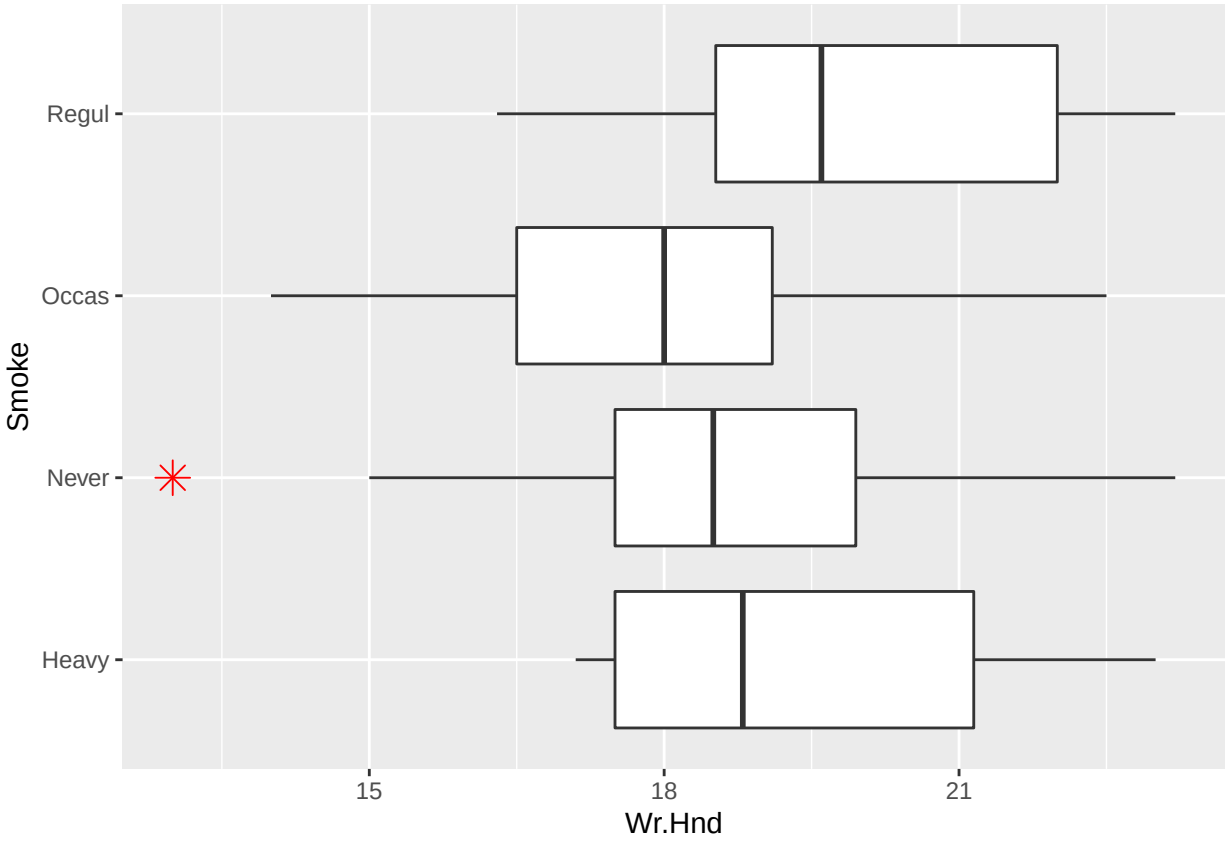
```
ggplot(newSurvey, aes(x=Age,color=Exer,fill=Exer)) +
facet_grid(.~Sex)+
geom_histogram(alpha=0.7)+
labs(title="Age Distribution", xlabel="Age Range", ylabel="Frequency")
```

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



12. Plot the box plot of writing hand span with respect to smoking habits of students.

```
ggplot(newSurvey, aes(x=Wr.Hnd, y=Smoke))+
geom_boxplot(outlier.colour="red", outlier.shape=8,outlier.size=4)
```


Regul
Occas
Never
Heavy
Smoke
15
18
21
Wr.Hnd