22BCE0476AmanChauhan.R

Batch1

2024-09-19

```
#Load the in-build dataset mtcars from R environment.
data("mtcars")
print(mtcars)
```

```
##                     mpg cyl  disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4          21.0   6 160.0 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag      21.0   6 160.0 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710         22.8   4 108.0  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive     21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout  18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
## Valiant            18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
## Duster 360         14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
## Merc 240D          24.4   4 146.7  62 3.69 3.190 20.00  1  0    4    2
## Merc 230           22.8   4 140.8  95 3.92 3.150 22.90  1  0    4    2
## Merc 280           19.2   6 167.6 123 3.92 3.440 18.30  1  0    4    4
## Merc 280C          17.8   6 167.6 123 3.92 3.440 18.90  1  0    4    4
## Merc 450SE         16.4   8 275.8 180 3.07 4.070 17.40  0  0    3    3
## Merc 450SL         17.3   8 275.8 180 3.07 3.730 17.60  0  0    3    3
## Merc 450SLC        15.2   8 275.8 180 3.07 3.780 18.00  0  0    3    3
## Cadillac Fleetwood 10.4   8 472.0 205 2.93 5.250 17.98  0  0    3    4
## Lincoln Continental 10.4  8 460.0 215 3.00 5.424 17.82  0  0    3    4
## Chrysler Imperial  14.7   8 440.0 230 3.23 5.345 17.42  0  0    3    4
## Fiat 128           32.4   4  78.7  66 4.08 2.200 19.47  1  1    4    1
## Honda Civic        30.4   4  75.7  52 4.93 1.615 18.52  1  1    4    2
## Toyota Corolla     33.9   4  71.1  65 4.22 1.835 19.90  1  1    4    1
## Toyota Corona      21.5   4 120.1  97 3.70 2.465 20.01  1  0    3    1
## Dodge Challenger   15.5   8 318.0 150 2.76 3.520 16.87  0  0    3    2
## AMC Javelin        15.2   8 304.0 150 3.15 3.435 17.30  0  0    3    2
## Camaro Z28         13.3   8 350.0 245 3.73 3.840 15.41  0  0    3    4
## Pontiac Firebird   19.2   8 400.0 175 3.08 3.845 17.05  0  0    3    2
## Fiat X1-9          27.3   4  79.0  66 4.08 1.935 18.90  1  1    4    1
## Porsche 914-2      26.0   4 120.3  91 4.43 2.140 16.70  0  1    5    2
## Lotus Europa       30.4   4  95.1 113 3.77 1.513 16.90  1  1    5    2
## Ford Pantera L     15.8   8 351.0 264 4.22 3.170 14.50  0  1    5    4
## Ferrari Dino       19.7   6 145.0 175 3.62 2.770 15.50  0  1    5    6
## Maserati Bora      15.0   8 301.0 335 3.54 3.570 14.60  0  1    5    8
## Volvo 142E         21.4   4 121.0 109 4.11 2.780 18.60  1  1    4    2
```

```
#Install the required packages and load in R. (psych - For statistical function, moments - Sk
ewness and kurtosis, ggplot..)
#install.packages("psych")
#install.packages("moments")
#install.packages("ggplot2")
#install.packages("dplyr")
library(psych)
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

```
##
## Attaching package: 'ggplot2'
```

```
## The following objects are masked from 'package:psych':
##
##     %+%, alpha
```

```
library(moments)
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.2.3
```

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

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

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

```
#Apply all the possibilities of 1D Statistical Data analysis.
#1.A measure of Central tendency (All possibilities of mean, all possibilities of median - Qu
antile, deciles (ntile), percentiles)
mean_arithmetic<-mean(mtcars$mpg)
mean_arithmetic
```

```
## [1] 20.09062
```

```
mean_harmonic<-psych::harmonic.mean(mtcars$mpg)
mean_harmonic
```

```
## [1] 18.44092
```

```
mean_geometric<-psych::geometric.mean(mtcars$mpg)
mean_geometric
```

```
## [1] 19.25006
```

```
median_value<-median(mtcars$mpg)
median_value
```

```
## [1] 19.2
```

```
quantiles<-quantile(mtcars$mpg,probs = seq(0,1,0.25))
quantiles
```

```
##     0%    25%    50%    75%   100%
## 10.400 15.425 19.200 22.800 33.900
```

```
deciles<-quantile(mtcars$mpg,probs = seq(0,1,0.1))
deciles
```

```
##    0%   10%   20%   30%   40%   50%   60%   70%   80%   90%  100%
## 10.40 14.34 15.20 15.98 17.92 19.20 21.00 21.47 24.08 30.09 33.90
```

```
percentiles<-quantile(mtcars$mpg,probs = seq(0,1,0.01))
percentiles
```

```
##     0%     1%     2%     3%     4%     5%     6%     7%     8%     9%    10%
## 10.400 10.400 10.400 10.400 11.096 11.995 12.894 13.470 13.780 14.090 14.340
##    11%    12%    13%    14%    15%    16%    17%    18%    19%    20%    21%
## 14.464 14.588 14.709 14.802 14.895 14.988 15.054 15.116 15.178 15.200 15.200
##    22%    23%    24%    25%    26%    27%    28%    29%    30%    31%    32%
## 15.200 15.239 15.332 15.425 15.518 15.611 15.704 15.797 15.980 16.166 16.352
##    33%    34%    35%    36%    37%    38%    39%    40%    41%    42%    43%
## 16.607 16.886 17.165 17.380 17.535 17.690 17.827 17.920 18.013 18.112 18.298
##    44%    45%    46%    47%    48%    49%    50%    51%    52%    53%    54%
## 18.484 18.670 18.830 18.985 19.140 19.200 19.200 19.200 19.260 19.415 19.570
##    55%    56%    57%    58%    59%    60%    61%    62%    63%    64%    65%
## 19.765 20.168 20.571 20.974 21.000 21.000 21.000 21.088 21.212 21.336 21.400
##    66%    67%    68%    69%    70%    71%    72%    73%    74%    75%    76%
## 21.400 21.400 21.408 21.439 21.470 21.513 21.916 22.319 22.722 22.800 22.800
##    77%    78%    79%    80%    81%    82%    83%    84%    85%    86%    87%
## 22.800 23.088 23.584 24.080 24.576 25.072 25.568 26.052 26.455 26.858 27.261
##    88%    89%    90%    91%    92%    93%    94%    95%    96%    97%    98%
## 28.168 29.129 30.090 30.400 30.400 30.400 30.680 31.300 31.920 32.505 32.970
##    99%   100%
## 33.435 33.900
```

```
#2.Measure of Dispersions (Range, IQR, Interdecile range, Deviation (mean and Standard deviat
ion), Skewness and Kurtosis)
range_value<-range(mtcars$mpg)
range
```

```
## function (..., na.rm = FALSE)  .Primitive("range")
```

```
IQR_value<-IQR(mtcars$mpg)
IQR_value
```

```
## [1] 7.375
```

```
interdecile_range<-quantile(mtcars$mpg,0.9)
interdecile_range
```

```
##   90%
## 30.09
```

```
mean_deviation<-mean(abs(mtcars$mpg-mean(mtcars$mpg)))
mean_deviation
```

```
## [1] 4.714453
```

```
sd_value<-sd(mtcars$mpg)
sd_value
```

```
## [1] 6.026948
```

```
skewness_value<-moments::skewness(mtcars$mpg)
skewness_value
```

```
## [1] 0.6404399
```

```
kurtosis_value<-moments::kurtosis(mtcars$mpg)
kurtosis_value
```

```
## [1] 2.799467
```

```
#3.Frequency Distribution with necessary plots ( Frequency Distribution, histogram, Relative frequency distribution and cumulative frequency distribution)
hist_data<-hist(mtcars$mpg,breaks=10,plot=FALSE)
hist_data
```

```
## $breaks
##  [1] 10 12 14 16 18 20 22 24 26 28 30 32 34
##
## $counts
##  [1] 2 1 7 3 5 5 2 2 1 0 2 2
##
## $density
##  [1] 0.031250 0.015625 0.109375 0.046875 0.078125 0.078125 0.031250 0.031250
##  [9] 0.015625 0.000000 0.031250 0.031250
##
## $mids
##  [1] 11 13 15 17 19 21 23 25 27 29 31 33
##
## $xname
## [1] "mtcars$mpg"
##
## $equidist
## [1] TRUE
##
## attr(,"class")
## [1] "histogram"
```

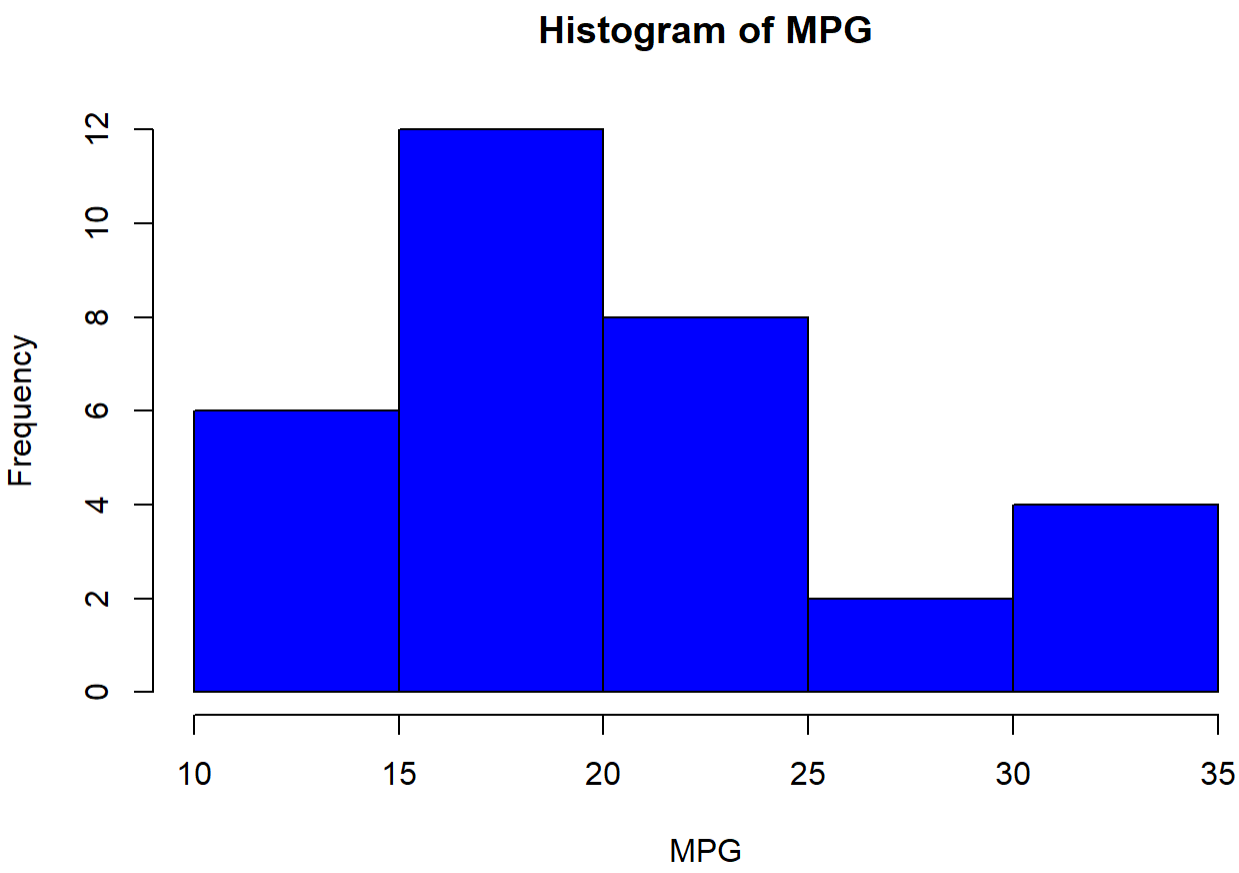
```
relative_frequency<-hist_data$counts/sum(hist_data$counts)
relative_frequency
```

```
##  [1] 0.06250 0.03125 0.21875 0.09375 0.15625 0.15625 0.06250 0.06250 0.03125
## [10] 0.00000 0.06250 0.06250
```

```
cumulative_frequency<-cumsum(relative_frequency)
cumulative_frequency
```

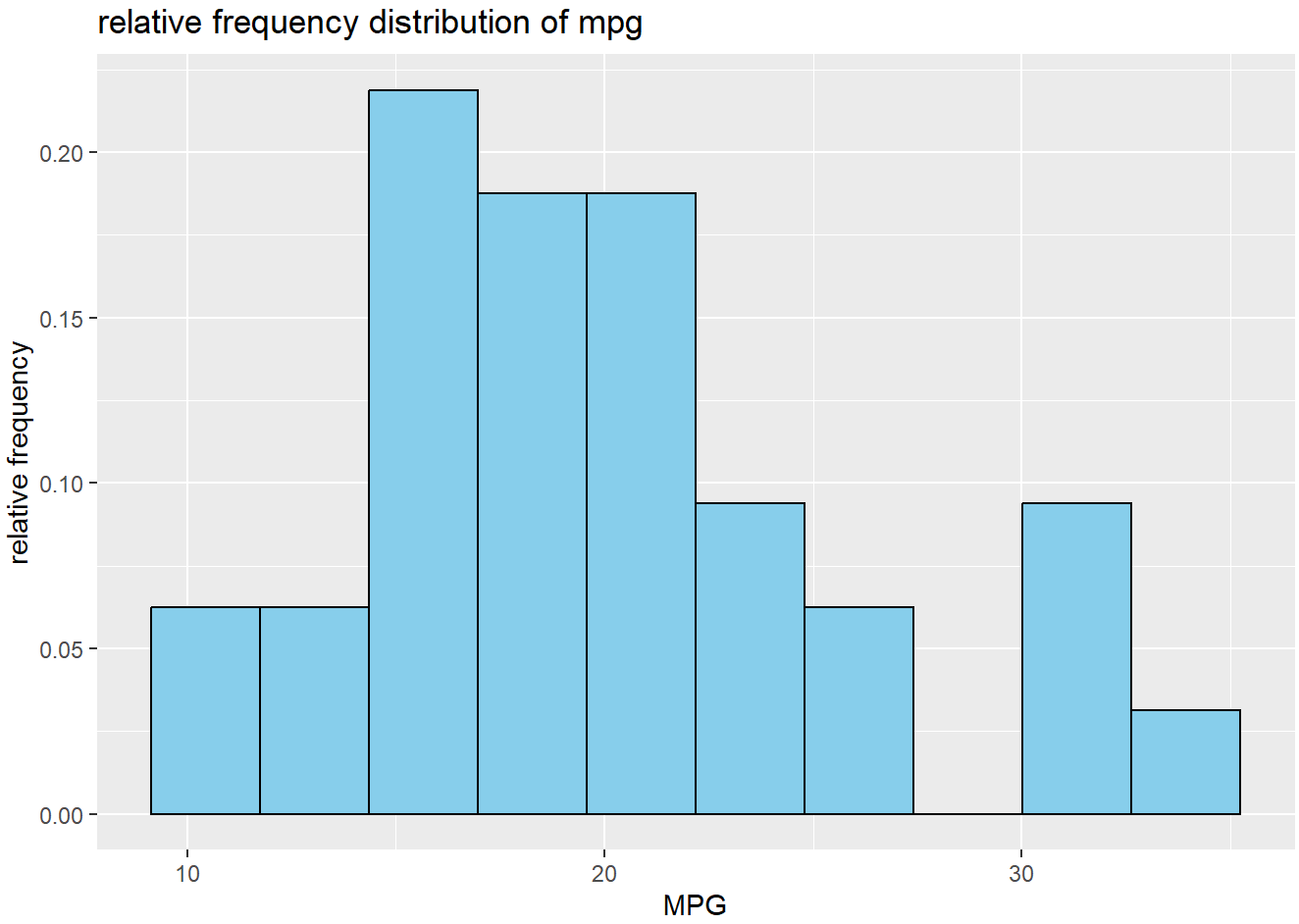
```
##  [1] 0.06250 0.09375 0.31250 0.40625 0.56250 0.71875 0.78125 0.84375 0.87500
## [10] 0.87500 0.93750 1.00000
```

```
hist(mtcars$mpg,main="Histogram of MPG",xlab = "MPG",ylab = "Frequency",col="blue")
```

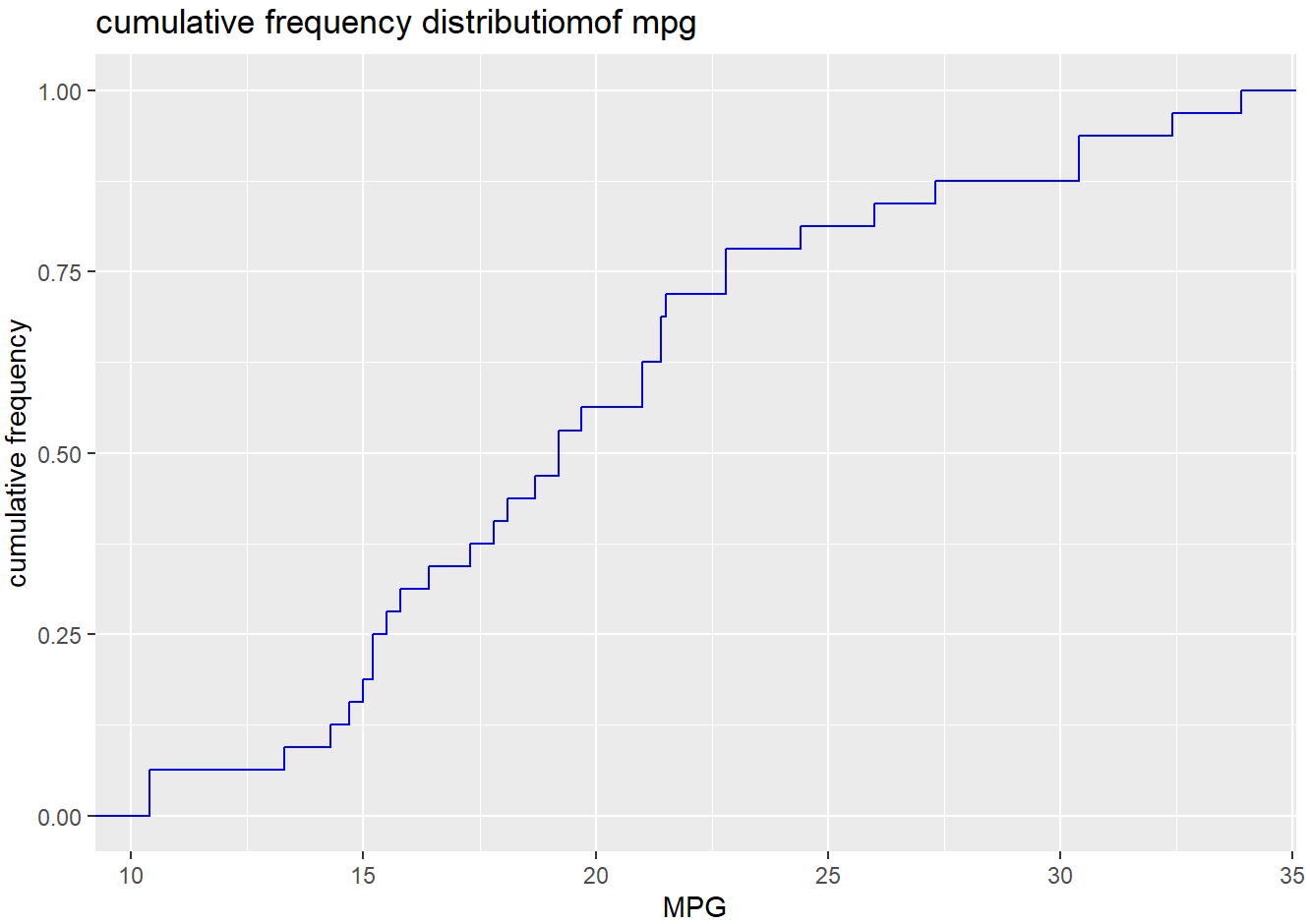



```
ggplot(mtcars,aes(x=mpg))+
  geom_histogram(aes(y=..count../sum(..count..)),bins = 10,fill="skyblue",col="black")+
  labs(title="relative frequency distribution of mpg",y="relative frequency",x="MPG")
```

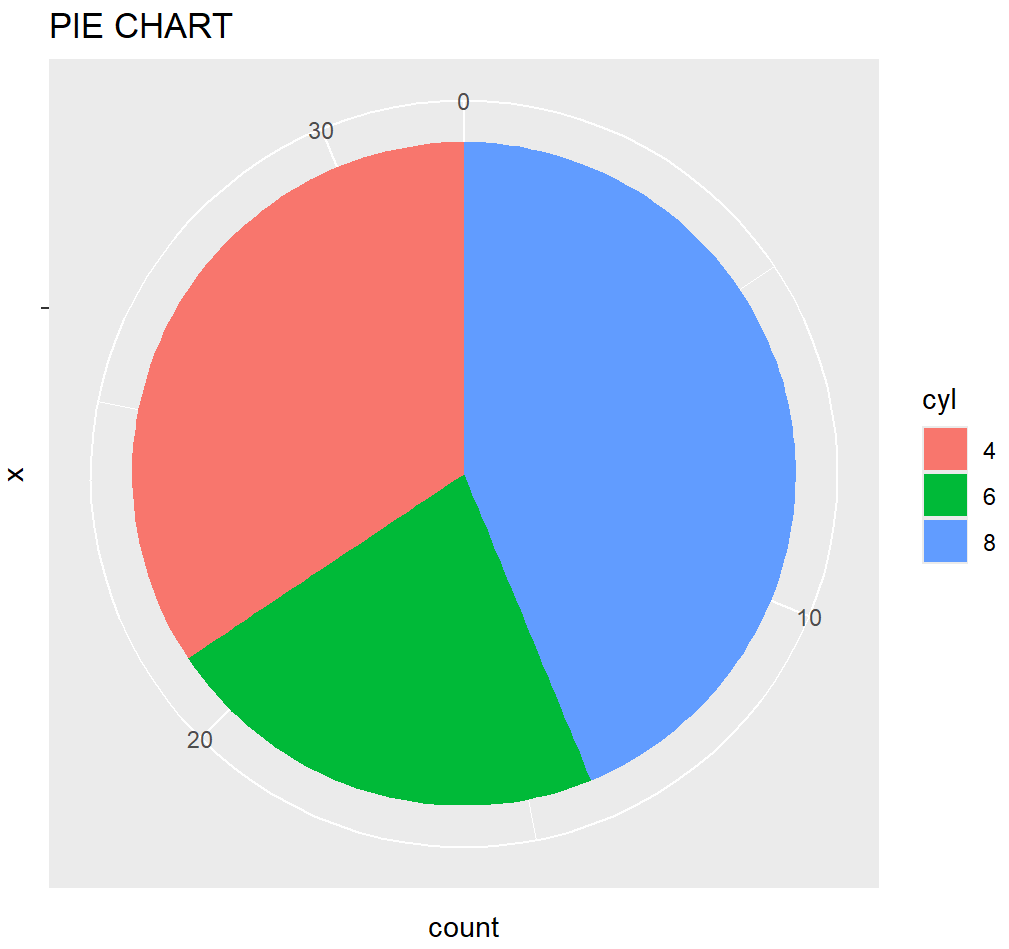
```
## Warning: The dot-dot notation (`..count..`) was deprecated in ggplot2 3.4.0.
## ℹ Please use `after_stat(count)` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



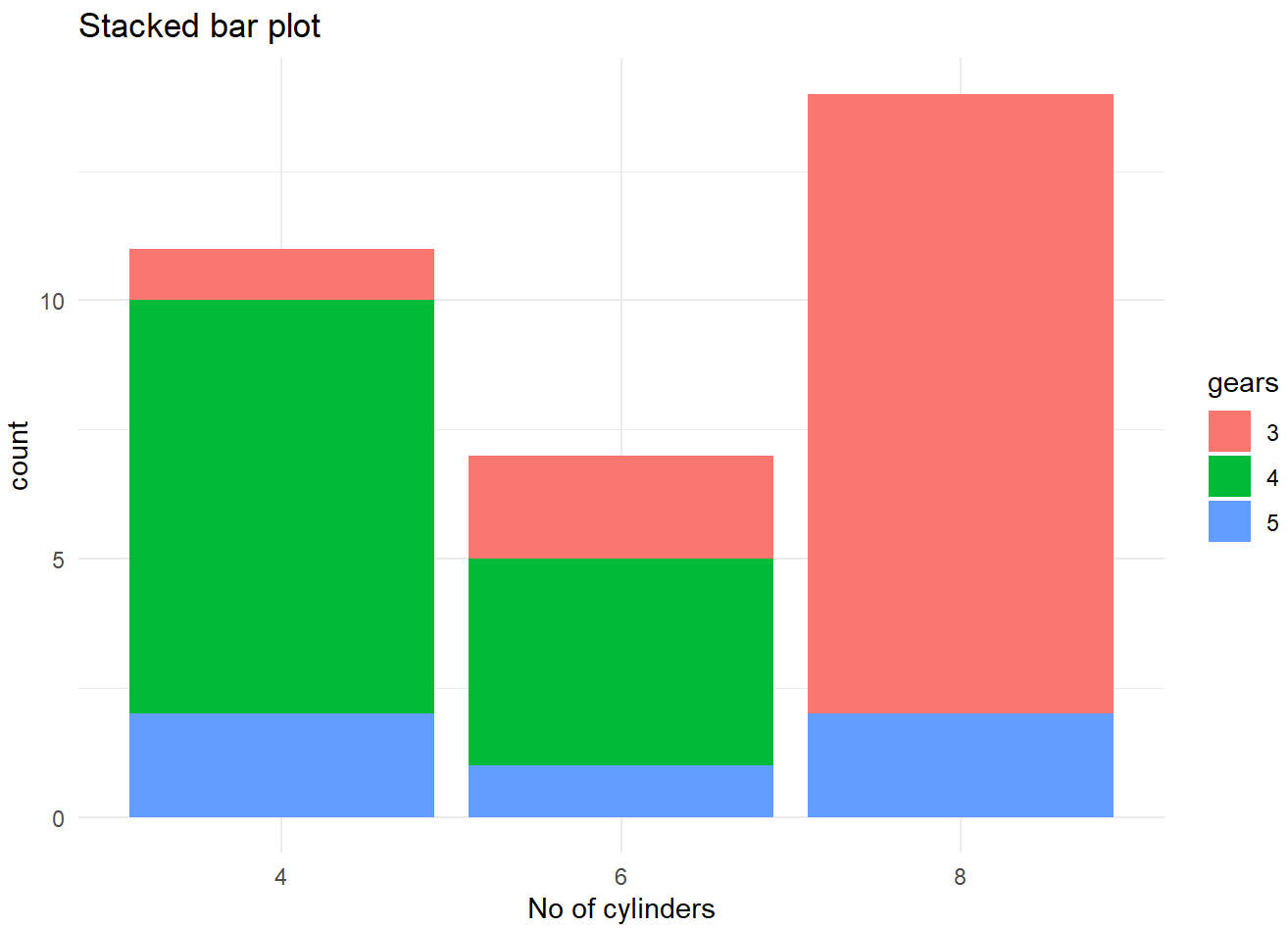

```
ggplot(mtcars,aes(x=mpg))+
  stat_ecdf(geom="step",col="blue")+
  labs(title="cumulative frequency distributiomof mpg",y="cumulative frequency",x="MPG")
```




```
#From the categorical variable ( Pie plot and Stacked bar plot)
#pie chart
mtcars$cyl<-as.factor(mtcars$cyl)
ggplot(mtcars,aes(x="",fill=cyl))+geom_bar(width=1)+coord_polar("y")+
  labs(title = "PIE CHART")
```




```
#Stacked bar plot
ggplot(mtcars,aes(x=as.factor(cyl),fill=factor(gear)))+
  geom_bar(position = "stack")+
  labs(title = "Stacked bar plot",x="No of cylinders",fill="gears")+theme_minimal()
```


Stacked bar plot
count
10
5
0
4
6
8
No of cylinders
gears
3
4
5