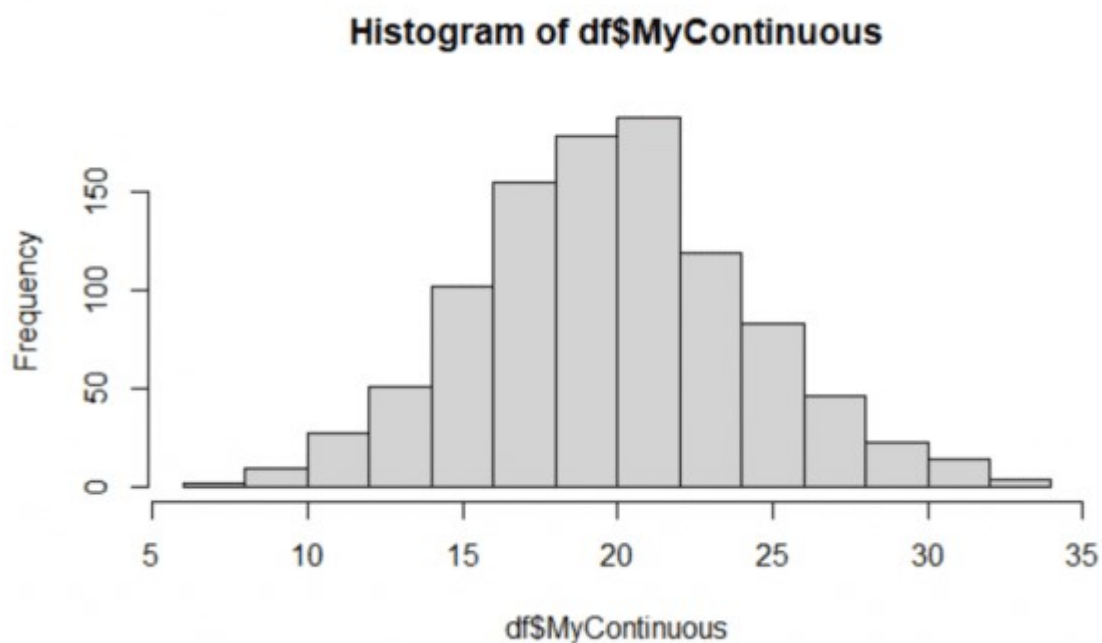


A very common task in data processing is the transformation of the numeric variables (continuous, discrete etc) to categorical by creating bins. For example, is quite often to convert the **age** to the **age group**. Let's see how we can easily do that in R.

We will consider a random variable from the Poisson distribution with parameter $\lambda=20$

```
library(dplyr)
# Generate 1000 observations from the Poisson distribution
# with lambda equal to 20
df<-data.frame(MyContinuous = rpois(1000,20))

# get the histogram
hist(df$MyContinuous)
```



Create specific Bins

Let's say that you want to create the following bins:

- **Bin 1: (-inf, 15]**
- **Bin 2: (15,25]**
- **Bin 3: (25, inf)**

We can easily do that using the `cut` command. Let's start:

```
df<-df%>%mutate(MySpecificBins = cut(MyContinuous, breaks =
c(-Inf,15,25,Inf)))
head(df,10)
```



```
head(df, 10)
```

```
> head(df, 10)
```

	MyContinuous	MySpecificBins	MyQuantileBins
1	24	(15, 25]	(23, 37]
2	24	(15, 25]	(23, 37]
3	18	(15, 25]	(17, 20]
4	24	(15, 25]	(23, 37]
5	23	(15, 25]	(20, 23]
6	22	(15, 25]	(20, 23]
7	15	(-Inf, 15]	(9, 17]
8	36	(25, Inf]	(23, 37]
9	15	(-Inf, 15]	(9, 17]
10	13	(-Inf, 15]	(9, 17]

We can check the `MyQuantileBins` if contain the same number of observations, and also to look at the ranges:

```
df%>%group_by(MyQuantileBins)%>%count()
```

```
> df%>%group_by(MyQuantileBins)%>%count()
```

A tibble: 4 x 2

Groups: MyQuantileBins [4]

	MyQuantileBins	n
	<fct>	<int>
1	[9, 17]	274
2	(17, 20]	259
3	(20, 23]	247
4	(23, 37]	220

Notice that in case that you want to split your continuous variable into bins of equal size you can also use the `ntile` function of the `dplyr` package, but it does not create labels of the bins based on the ranges.