

Alopecia

Qiang Liu

2022-09-21

First include basic settings and related packages.

```
> library(tidyverse)
```

```
## Warning:  'tidyverse' R 4.2.1
```

```
## Warning:  'ggplot2' R 4.2.1
```

```
## Warning:  'tibble' R 4.2.1
```

```
## Warning:  'dplyr' R 4.2.1
```

```
## Warning:  'forcats' R 4.2.1
```

```
> library(skimr)
```

```
## Warning:  'skimr' R 4.2.1
```

```
> library(MASS)
```

```
## Warning:  'MASS' R 4.2.1
```

```
> library(openxlsx)
```

```
## Warning:  'openxlsx' R 4.2.1
```

```
> library(pander)
```

```
## Warning:  'pander' R 4.2.1
```

```
> library(ggplot2)
```

```
> library(Rcpp)
```

```
## Warning:  'Rcpp' R 4.2.1
```

```
> library(dplyr)
> library(viridis)
```

```
## Warning: 'viridis' R 4.2.1
```

```
> library(ggstatsplot)
```

```
## Warning: 'ggstatsplot' R 4.2.1
```

```
> library(plyr)
> library(hrbrthemes)
```

```
## Warning: 'hrbrthemes' R 4.2.1
```

Read and roughly view data ‘Alopecia’ in project’s working directory

```
> age_countrydata <- read.xlsx("Age and Country.xlsx")
> countrydata <- read.xlsx("Country.xlsx")
> gender_degreedata <- read.xlsx("Gender and Degree.xlsx")
> age_degreedata <- read.xlsx("Age and Degree.xlsx")
> skim(age_countrydata)
```

Table 1: Data summary

Name	age_countrydata
Number of rows	3047
Number of columns	4
Column type frequency:	
character	2
numeric	2
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
Country	0	1.00	5	6	0	6	0
Gender	49	0.98	1	1	0	2	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
Age	0	1	35.93	15.13	15	24	32	44	93	
Time	0	1	2016.55	3.36	2012	2012	2018	2019	2022	

```
> skim(countrydata)
```

Table 4: Data summary

Name	countrydata
Number of rows	19
Number of columns	2
Column type frequency:	
character	1
numeric	1
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
Country	0	1	3	9	0	19	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
Count	0	1	420.79	618.63	20	66	130	479.5	2495	

```
> skim(gender_degreedata)
```

Table 7: Data summary

Name	gender_degreedata
Number of rows	1290
Number of columns	4
Column type frequency:	
character	2
numeric	2
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
Country	0	1	3	8	0	6	0
Gender	0	1	1	1	0	2	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
HNS	460	0.64	3.14	1.44	1	2	3	4	9	
LS	830	0.36	1.53	0.81	0	1	1	2	3	

```
> skim(age_degreedata)
```

Table 10: Data summary

Name	age_degreedata
Number of rows	419
Number of columns	5
Column type frequency:	
character	2
numeric	3
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
Country	0	1	5	5	0	2	0
Gender	0	1	1	1	0	2	0

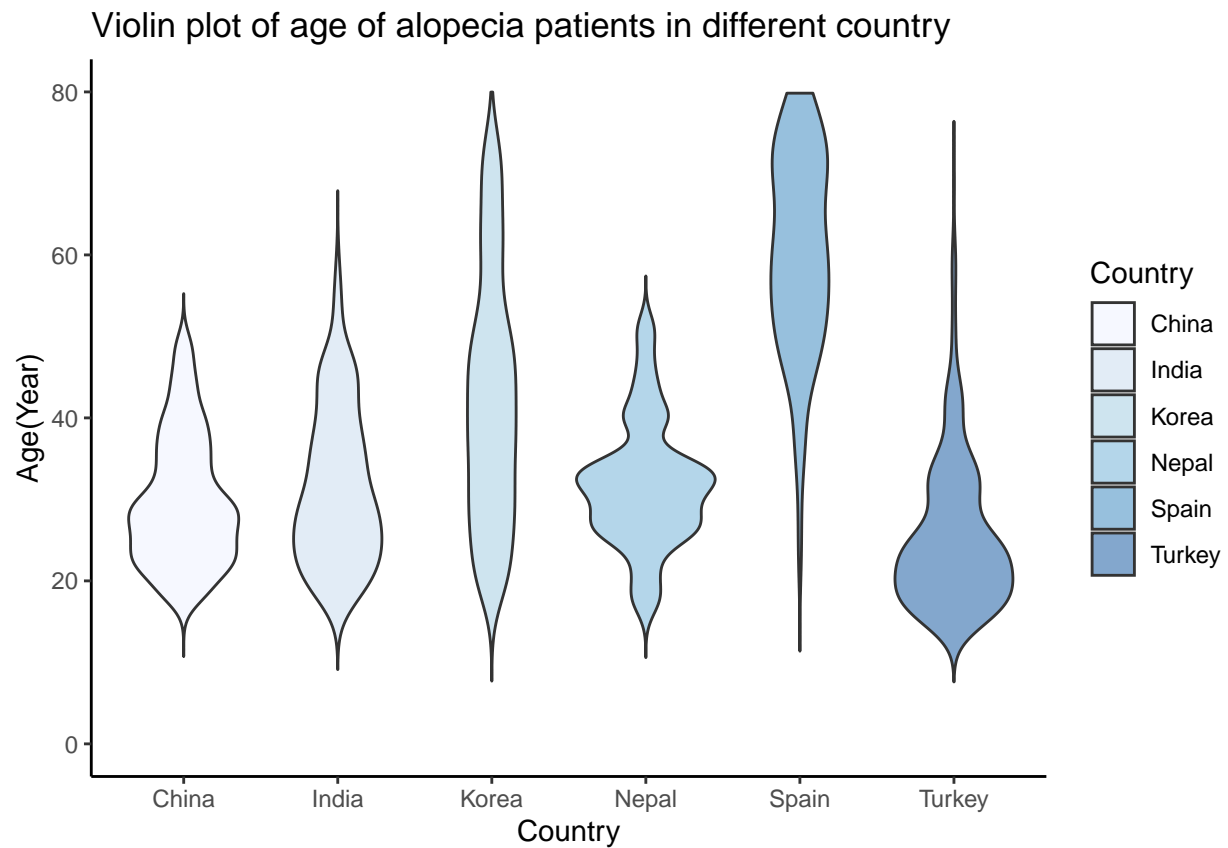
Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
Age	0	1.00	48.29	20.20	18	25.5	45.5	64	93	
HNS	53	0.87	3.39	1.51	1	2.0	3.0	4	7	
LS	366	0.13	0.85	1.15	0	0.0	0.0	2	3	

```
> age_countrydata%>%
+   ggplot(aes(x=Country,y=Age))+
+   geom_violin(aes(fill=Country),trim=FALSE,alpha=0.5)+
+   scale_fill_brewer()+
+   xlab("Country")+
+   ylab("Age(Year)")+
+   ylim(0,80)+
+   ggtitle("Violin plot of age of alopecia patients in different country")+
+   theme_classic()
```

```
## Warning: Removed 28 rows containing non-finite values (stat_ydensity).
```

```
## Warning: Removed 110 rows containing missing values (geom_violin).
```



```
> pdf("boxplot of country.pdf",width=8,height=8)
> ggbetweenstats(
+   data=age_countrydata,
+   x=Country,
+   y=Age,
+   title="Age distribution of alopecia patients in different country",
+   message=FALSE
+ )
```

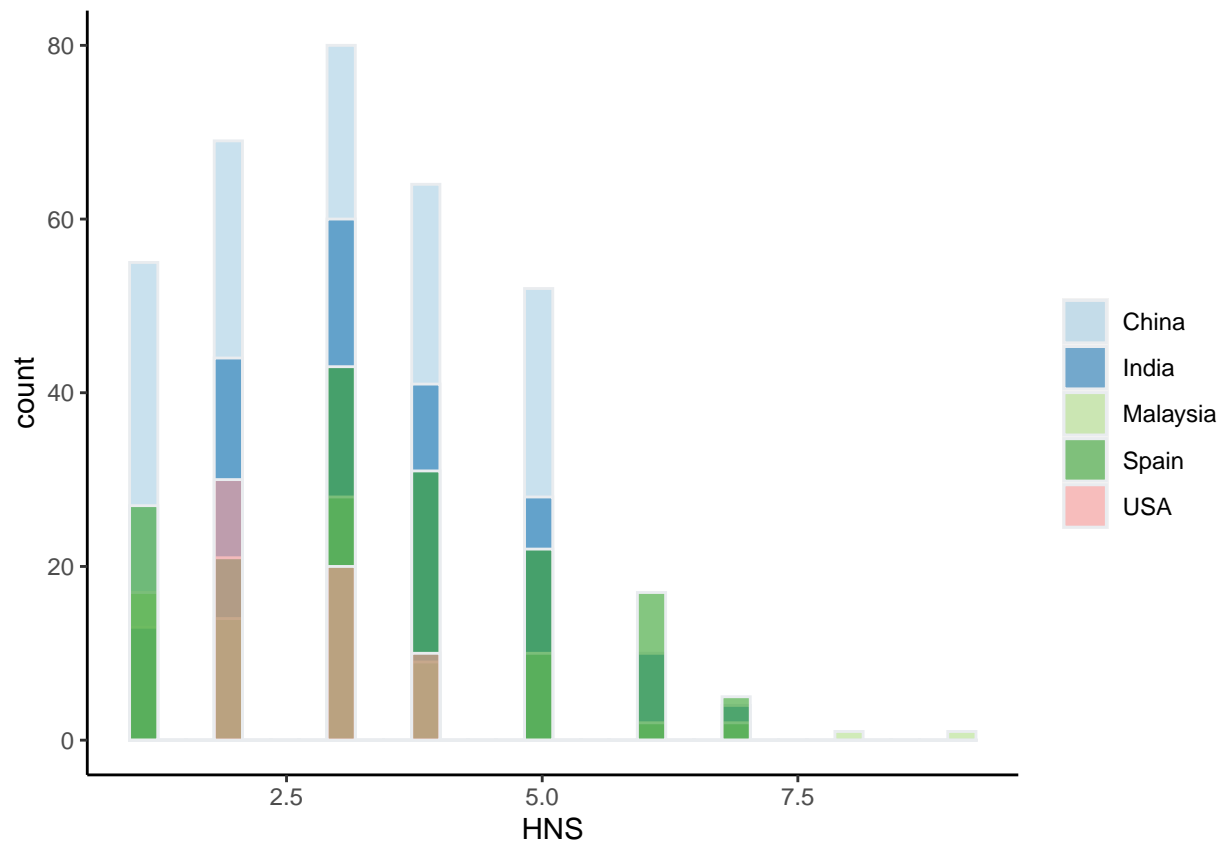
```
> pdf("boxplot of gender.pdf",width=8,height=8)
> ggbetweenstats(
+   data=age_countrydata,
+   x=Gender,
+   y=Age,
+   title="Age distribution of alopecia patients in different gender",
+   message=FALSE
+ )
```

```
> pdf("boxplot of time.pdf",width=8,height=8)
> ggbetweenstats(
+   data=age_countrydata,
+   x=Time,
+   y=Age,
+   title="Age distribution of alopecia patients in different Time",
+   message=FALSE
+ )
```

```
> male_degreedata <- gender_degreedata[grepl(pattern="M",gender_degreedata$Gender),]
> female_degreedata <- gender_degreedata[grepl(pattern="F",gender_degreedata$Gender),]
```

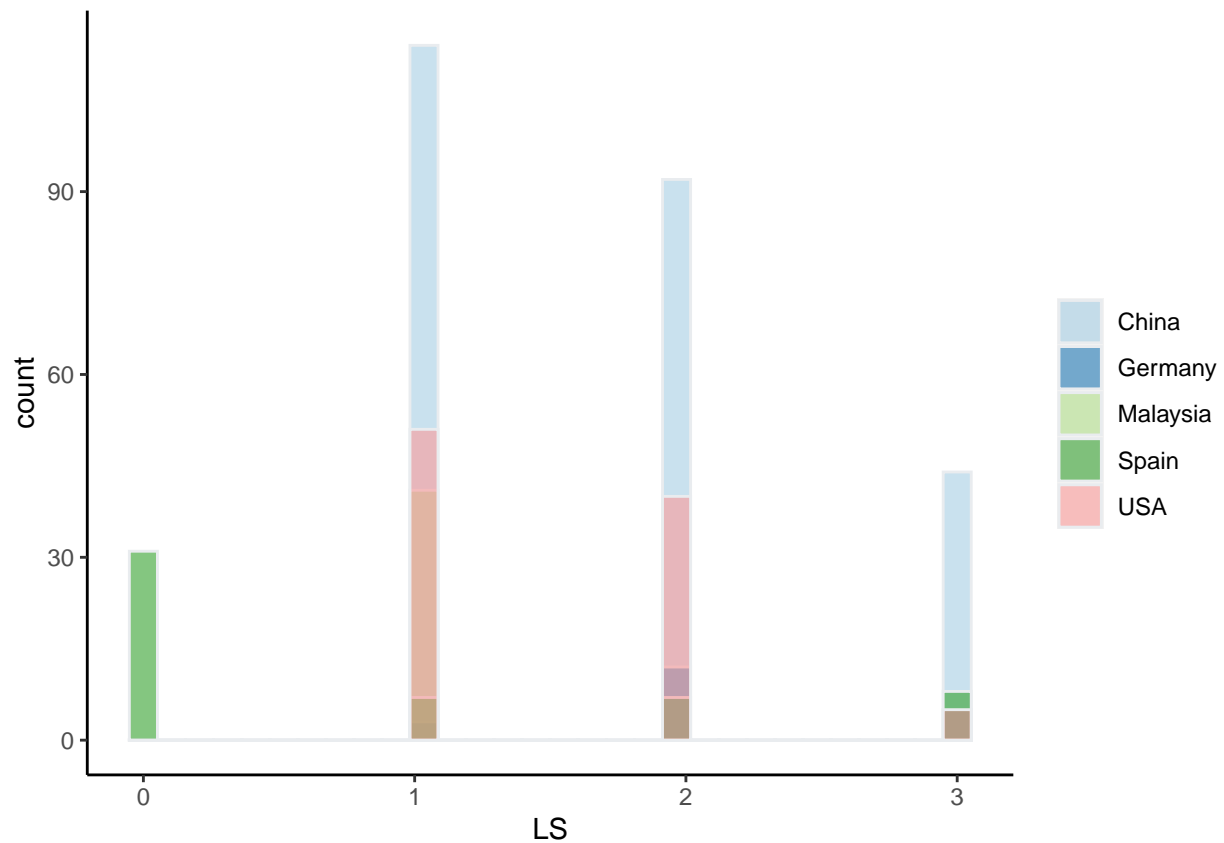
```
> male_degreedata %>%
+   ggplot(aes(x=HNS, fill=Country)) +
+     geom_histogram(color="#e9ecef", alpha=0.6, position = 'identity') +
+     scale_fill_brewer(palette = "Paired") +
+     labs(fill="") +
+     theme(panel.grid.major=element_blank(),panel.grid.minor=element_blank(),
+           panel.background=element_blank(),axis.line=element_line(colour="black"))
```

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



```
> female_degreedata %>%
+   ggplot(aes(x=LS, fill=Country)) +
+     geom_histogram(color="#e9ecef", alpha=0.6, position = 'identity') +
+     scale_fill_brewer(palette = "Paired") +
+     labs(fill="")+
+     theme(panel.grid.major=element_blank(),panel.grid.minor=element_blank(),
+           panel.background=element_blank(),axis.line=element_line(colour="black"))
```

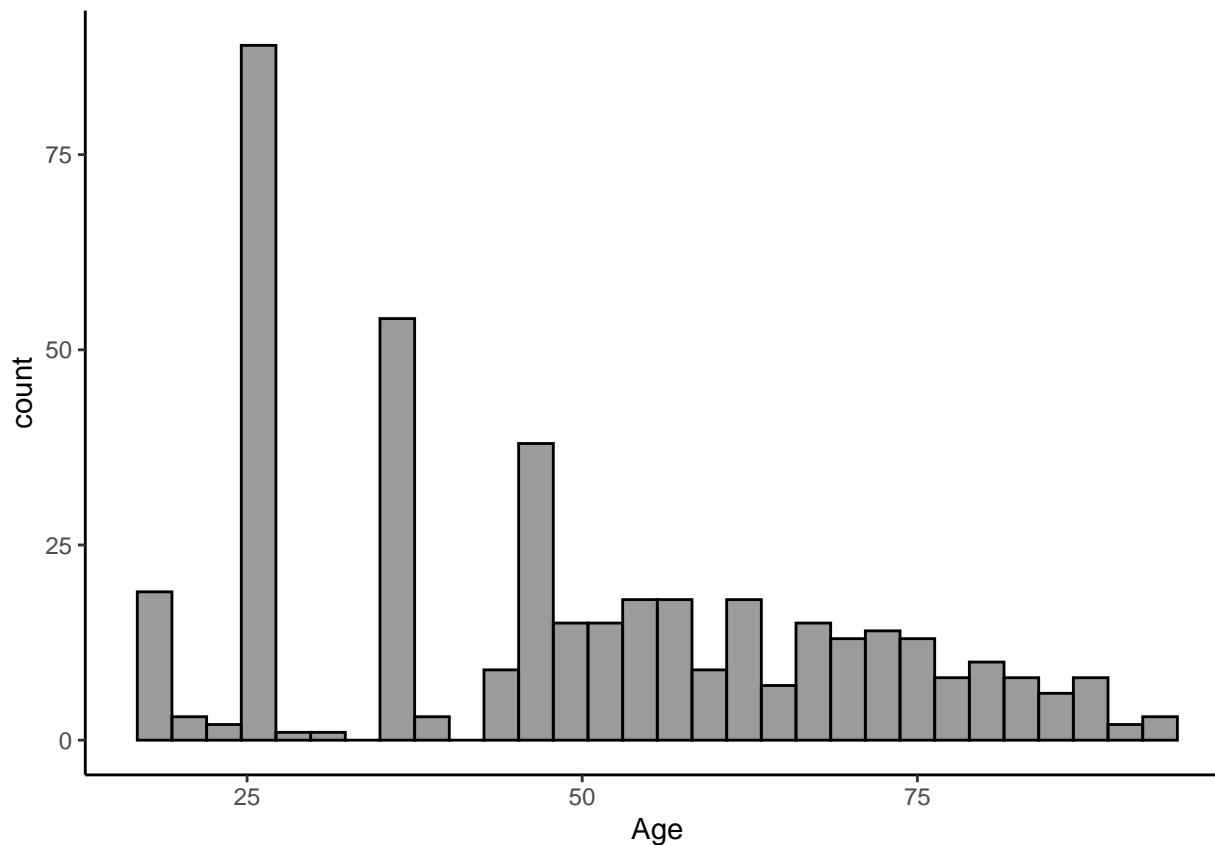
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.



```
> pdf("boxplot of female degree.pdf",width=8,height=8)
> ggbetweenstats(
+   data=female_degreedata,
+   x=Country,
+   y=LS,
+   title="LS distribution of alopecia patients in different Country",
+   message=FALSE
+ )
```

```
> age_degreedata %>%
+   ggplot(aes(x=Age)) +
+     geom_histogram(color="black", alpha=0.6, position = 'identity') +
+     scale_fill_continuous() +
+     labs(fill="")+
+     theme(panel.grid.major=element_blank(),panel.grid.minor=element_blank(),
+           panel.background=element_blank(),axis.line=element_line(colour="black"))
```

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

```
> male_agedata <- age_degreedata[grep(pattern="M",age_degreedata$Gender),]
```

```
> linear.model <- lm(HNS~Age,data=male_agedata)
> summary(linear.model)
```

```
##
## Call:
## lm(formula = HNS ~ Age, data = male_agedata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3406 -0.9068  0.0516  0.8935  3.6030
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.281867   0.193280  11.806 < 2e-16 ***
## Age          0.024508   0.003928   6.239 1.23e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.433 on 364 degrees of freedom
## Multiple R-squared:  0.0966, Adjusted R-squared:  0.09411
## F-statistic: 38.92 on 1 and 364 DF, p-value: 1.227e-09
```

```
> stepwiseSelection <- stepAIC(linear.model,direction="both",trace=FALSE,k=log(NROW(male_agedata)))
> summary(stepwiseSelection)
```

```
##
## Call:
## lm(formula = HNS ~ Age, data = male_agedata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3406 -0.9068  0.0516  0.8935  3.6030
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.281867   0.193280  11.806 < 2e-16 ***
## Age          0.024508   0.003928   6.239 1.23e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.433 on 364 degrees of freedom
## Multiple R-squared:  0.0966, Adjusted R-squared:  0.09411
## F-statistic: 38.92 on 1 and 364 DF, p-value: 1.227e-09
```

```
> pander::pander(stepwiseSelection)
```

Table 13: Fitting linear model: HNS ~ Age

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.282	0.1933	11.81	1.871e-27
Age	0.02451	0.003928	6.239	1.227e-09

```
> #Ssadasdasdasdasdsadasdsa
> #sadsadsadasdsadsadsad
> #sadasdasdasdasdassadasda
> #sadasgregerhgefwerfasfd
> #sdzgayhrhava bgfdhd
> #dgdsfhsjhrthjerherh
> #sdgsdghsdhdfhhfdgsgsd
```

$$\widehat{HNS}(male) = 2.282 + 0.02451 * Age$$

```
> pdf("Cor of Age and Degree of male androgenetic alopecia (HNS).pdf",width=9,height=7.5)
> ggscatterstats(
+   data=male_agedata,
+   x=Age,
+   y=HNS,
+   type="p",
+   bin=30,
+   conf.level=0.99,
+   messages=TRUE
+ )
```

```
## Registered S3 method overwritten by 'ggside':
##   method from
##   +.gg      ggplot2
```

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

```
> dev.off
```

```
## function (which = dev.cur())
## {
##   if (which == 1)
##     stop("cannot shut down device 1 (the null device)")
##   .External(C_devoff, as.integer(which))
##   dev.cur()
## }
## <bytecode: 0x000001ddfdfa5ef0>
## <environment: namespace:grDevices>
```
