R Notebook

```r
library(psych)
```

```r
data <- read.csv('data.csv')
```

```r
#View(data)
```

```r
data$HG_NHG<-as.factor(data$HG_NHG)
data$HG_NHG2<-as.factor(data$HG_NHG2)
```

```r
data$Status1<-as.factor(data$Status1)
```

```r
summary(data)
```

```
##        X            country              year_x          CL
##  Min.   :  0.00   Length:152         Min.   :2019   Min.   :1.000
##  1st Qu.: 38.75   Class :character   1st Qu.:2019   1st Qu.:2.000
##  Median : 77.50   Mode  :character   Median :2019   Median :3.000
##  Mean   : 77.40                      Mean   :2019   Mean   :3.349
##  3rd Qu.:115.25                      3rd Qu.:2019   3rd Qu.:5.000
##  Max.   :156.00                      Max.   :2019   Max.   :7.000
##        PR           Status           Status1     year_y        gdp_per
##  Min.   :1.000   Length:152         F :63   Min.   :2019   Min.   :   784.9
##  1st Qu.:1.750   Class :character   NF:89   1st Qu.:2019   1st Qu.:  5263.1
##  Median :3.000   Mode  :character           Median :2019   Median : 14844.8
##  Mean   :3.401                              Mean   :2019   Mean   : 22730.4
##  3rd Qu.:5.000                              3rd Qu.:2019   3rd Qu.: 33100.6
##  Max.   :7.000                              Max.   :2019   Max.   :124590.6
##     region           adminregion        incomeLevel          income
##  Length:152         Length:152         Length:152         Length:152
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##     variable        value         HG_NHG     status2            income2
##  Min.   :2019   Min.   :16.00   High:91   Length:152         Length:152
##  1st Qu.:2019   1st Qu.:29.75   Low :61   Class :character   Class :character
##  Median :2019   Median :40.00             Mode  :character   Mode  :character
##  Mean   :2019   Mean   :44.40
##  3rd Qu.:2019   3rd Qu.:56.00
##  Max.   :2019   Max.   :87.00
##  HG_NHG2
##  High: 49
##  Low :103
```

```
##
##
##
##
```

```
library(dplyr)
```

```
##
## 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
```

```
data %>% count(Status1, HG_NHG)
```

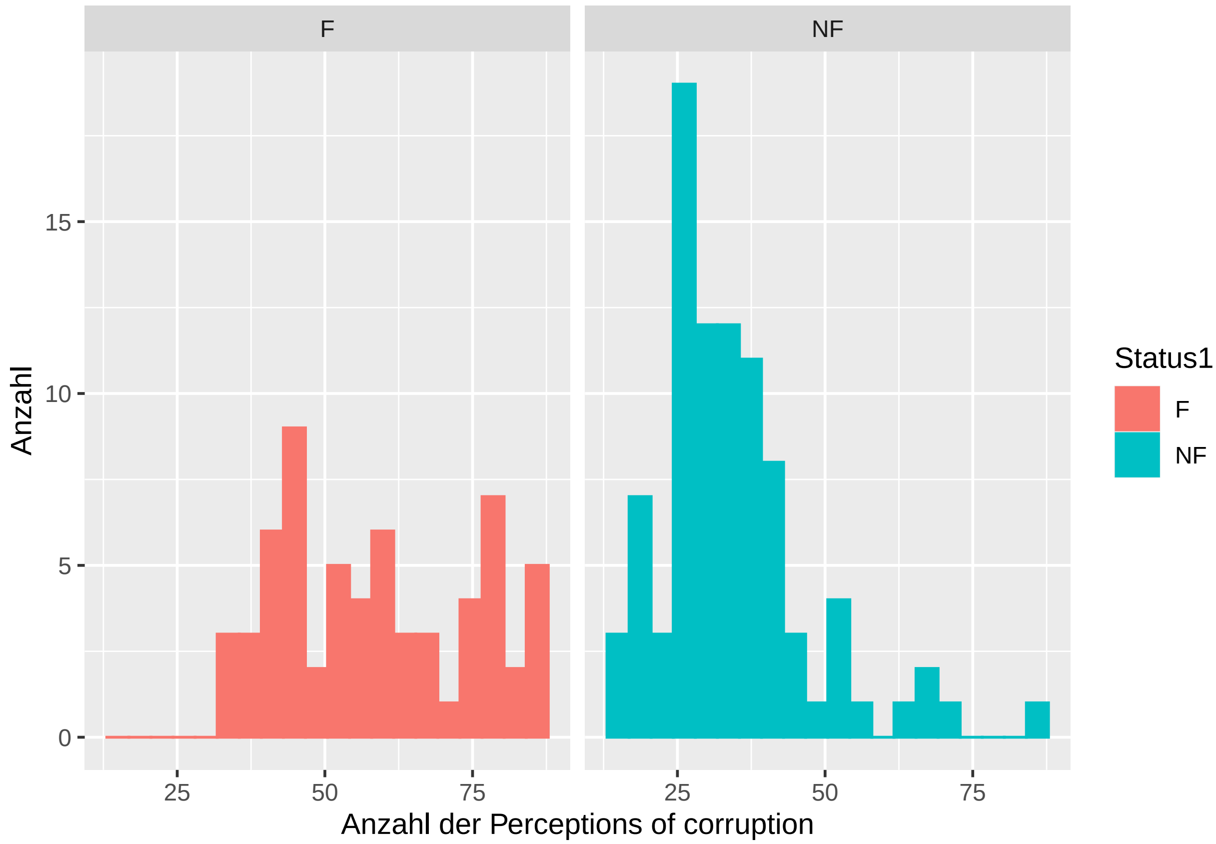
```
##   Status1 HG_NHG  n
## 1       F   High 53
## 2       F    Low 10
## 3      NF   High 38
## 4      NF    Low 51
```

```
library(ggplot2)
```

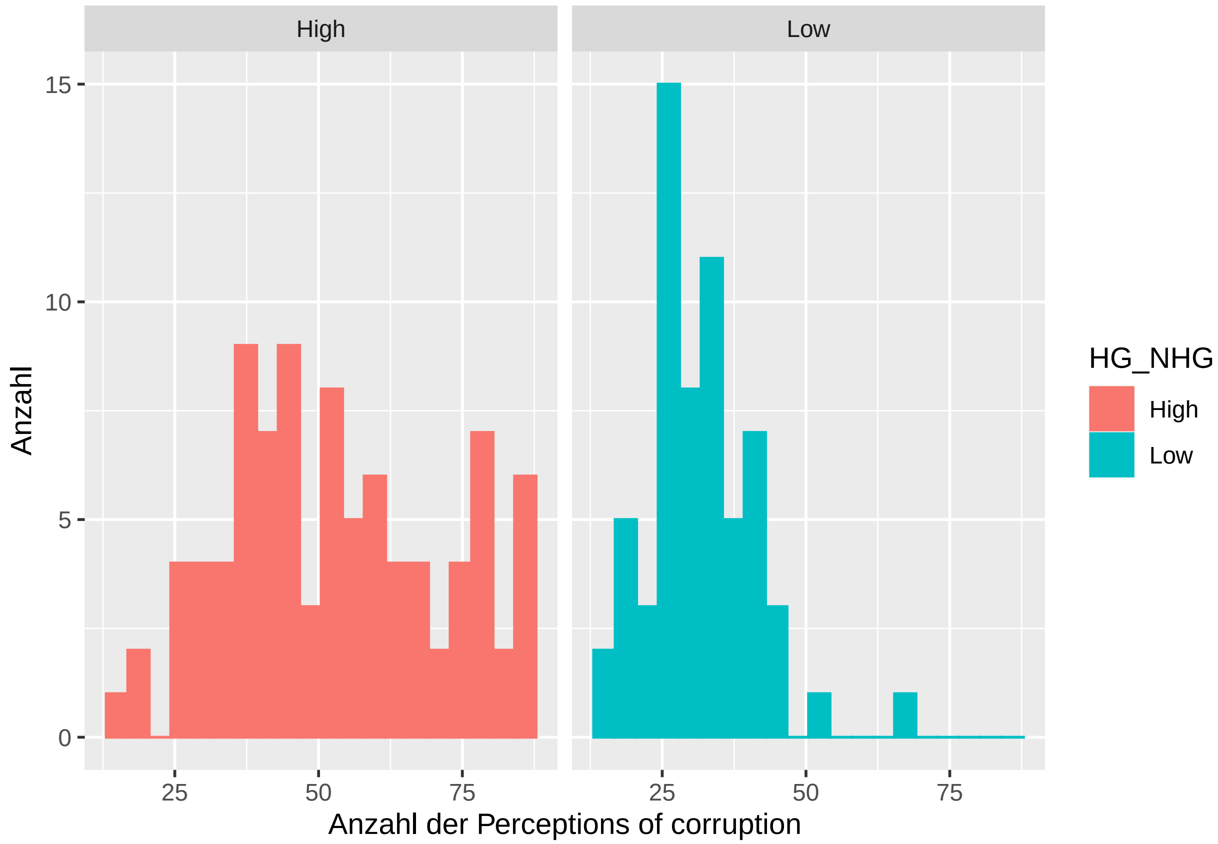
```
##
## Attaching package: 'ggplot2'

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

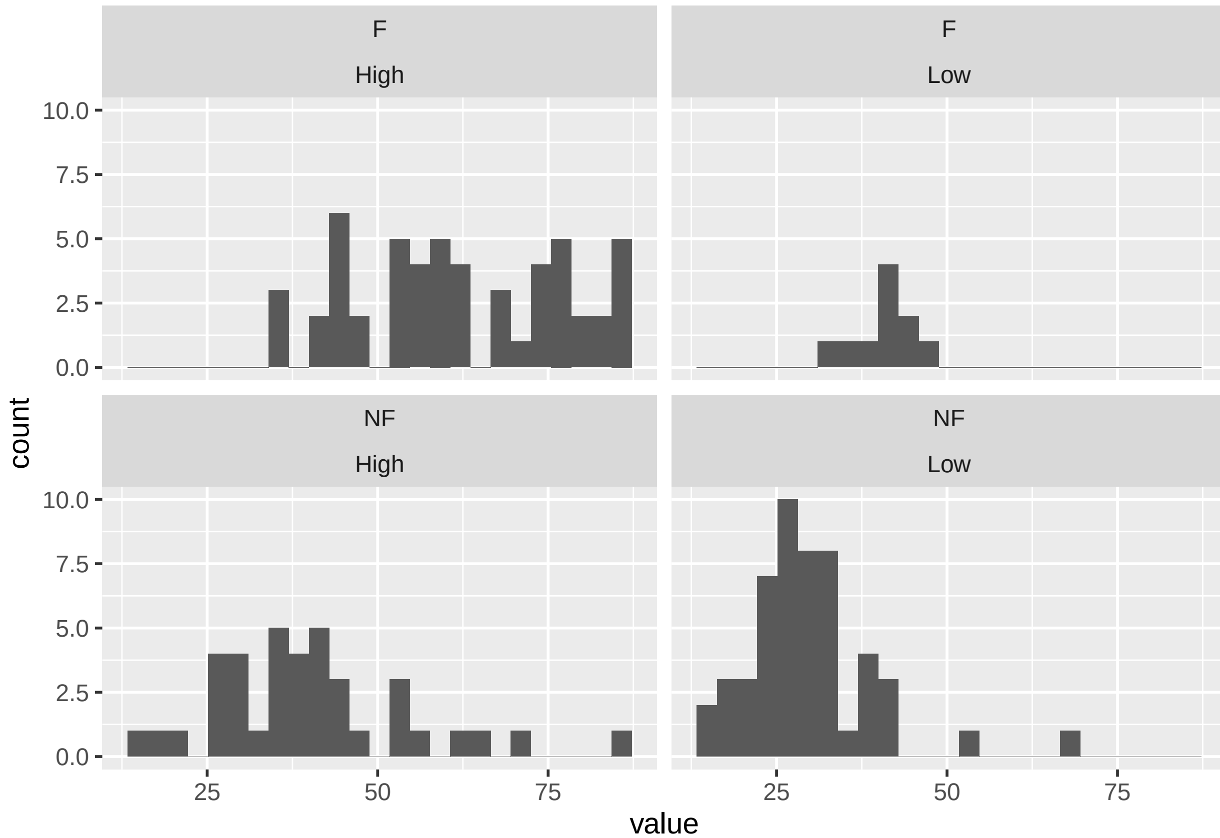
```
data %>%
  group_by(Status1) %>%
  ggplot(aes(value, color=Status1)) +
  geom_histogram(aes(fill = Status1), bins = 20) +
  facet_wrap(~Status1) +
  theme_grey()+
  labs(x= "Anzahl der Perceptions of corruption",y = "Anzahl" )
```



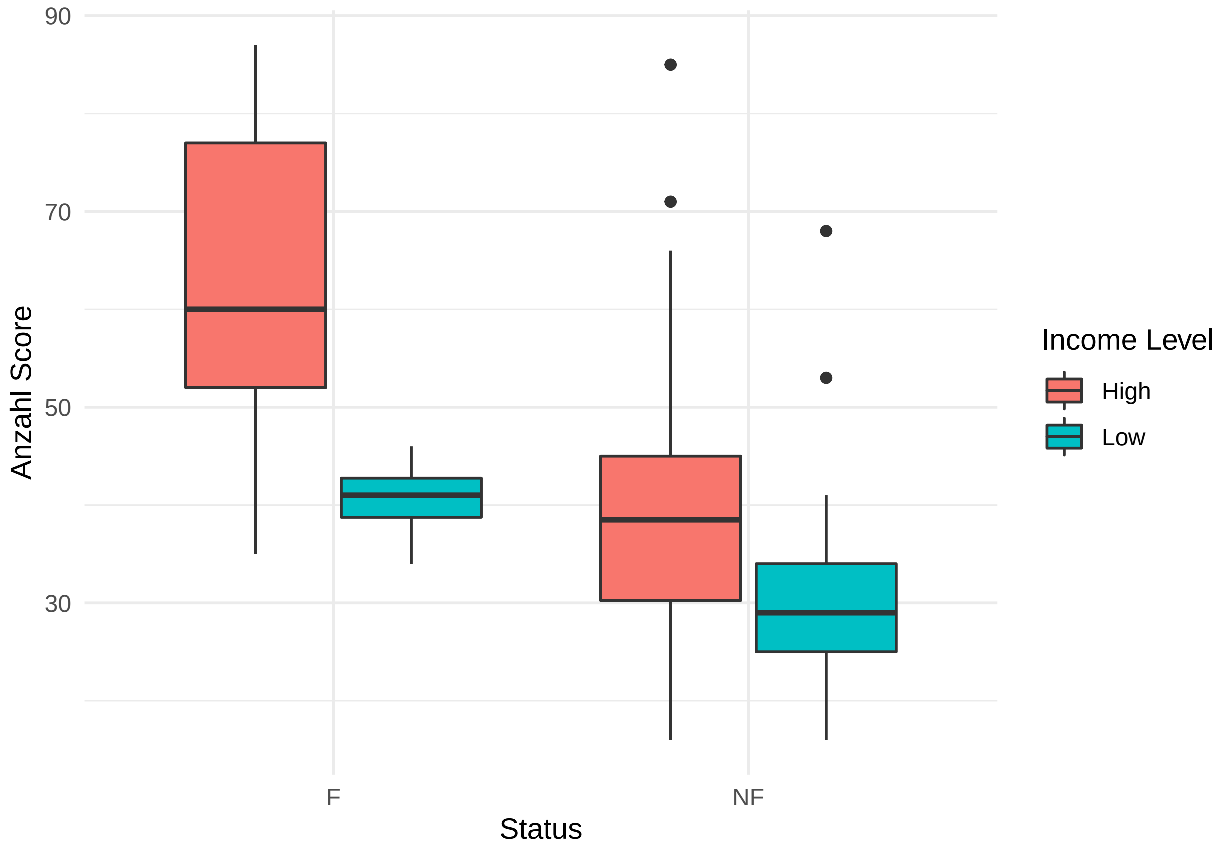

```
data %>%
  group_by(HG_NHG) %>%
  ggplot(aes(value, color=HG_NHG)) +
  geom_histogram(aes(fill = HG_NHG), bins = 20) +
  facet_wrap(~HG_NHG) +
  theme_grey()+
  labs(x= "Anzahl der Perceptions of corruption",y = "Anzahl" )
```




```
ggplot(data, aes(x=value))+
geom_histogram(bins = 25)+
theme_grey()+
facet_wrap(Status1~HG_NHG, ncol = 2)
```



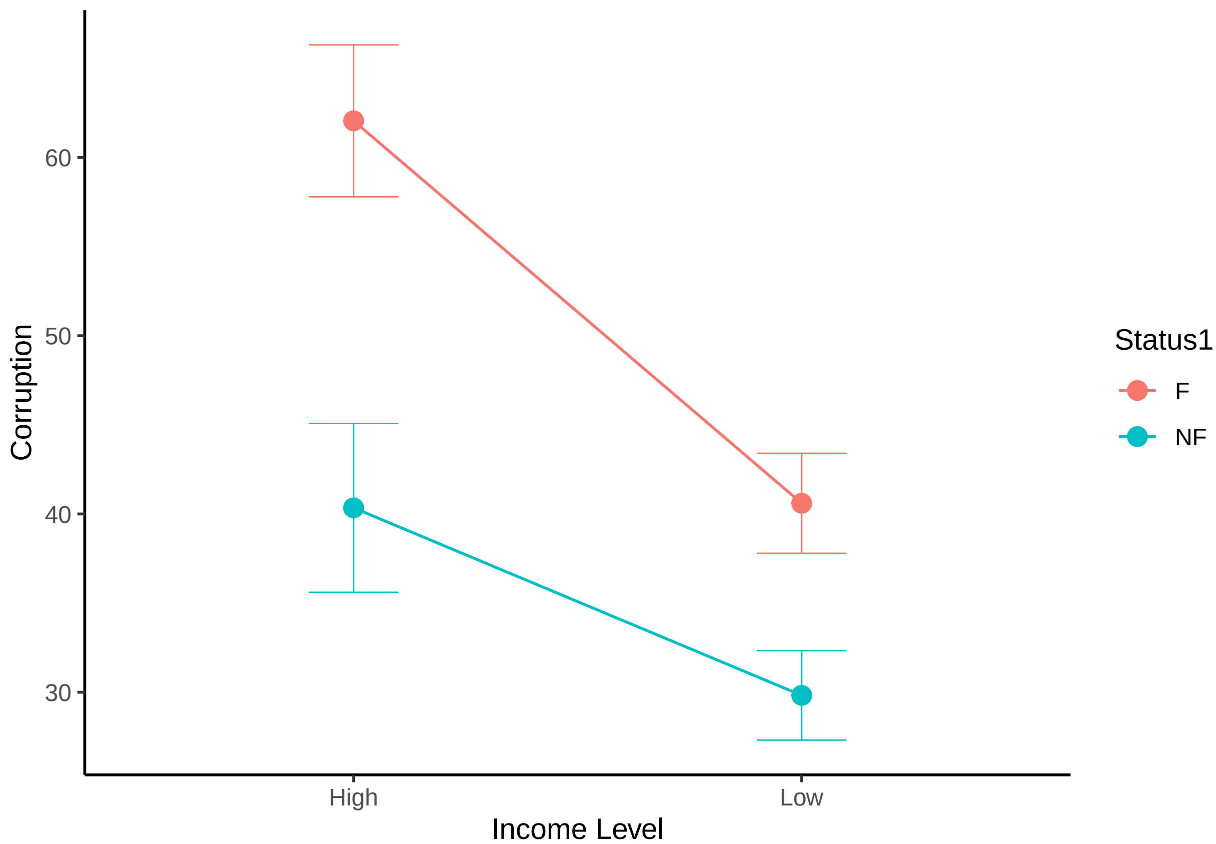

```
ggplot(data, aes(Status1, value, fill=factor(HG_NHG))) +
geom_boxplot() +
theme_minimal()+
labs(fill = "Income Level", x="Status", y="Anzahl Score")
```




```
ggplot(data, aes(x=HG_NHG, y=value, group=Status1, color = Status1))+
  stat_summary(fun.y = mean, geom="point", size=3)+
  stat_summary(fun.y = mean, geom="line")+
  stat_summary(fun.data = mean_cl_normal, geom="errorbar",width=.2, size=.25)+
  labs(x="Income Level", y="Corruption")+
  theme_classic()
```

```
## Warning: `fun.y` is deprecated. Use `fun` instead.

## Warning: `fun.y` is deprecated. Use `fun` instead.
```




```
library(car)
```

```
## Loading required package: carData

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##     recode

## The following object is masked from 'package:psych':
##
##     logit
```

```
leveneTest(value ~ Status1*HG_NHG, data = data, center = "mean")
```

```
## Levene's Test for Homogeneity of Variance (center = "mean")
##        Df F value    Pr(>F)
## group   3  9.7704 6.386e-06 ***
##       148
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Mit Welch-Korrektur: $p < 0.05$ => Ergebnis Signifikant –> Varianzen heterogen

Ohne Welch-Korrektur: $p > 0.05$ => Ergebnis nicht Signifikant –> Varianzen homogen –> H0 mit Annahme Var1=Var2=… -> Var_n wird angenommen

```
multiAnova1 <- lm(value ~ Status1*HG_NHG, data = data)
myAnova <- Anova(multiAnova1 , type = 3)
myAnova
```

```
## Anova Table (Type III tests)
##
## Response: value
##                 Sum Sq  Df  F value    Pr(>F)
## (Intercept)    204104   1 1246.531 < 2.2e-16 ***
## Status1         10436   1   63.734 3.641e-13 ***
## HG_NHG           3873   1   23.654 2.916e-06 ***
## Status1:HG_NHG    726   1    4.434   0.03692 *
## Residuals       24233 148
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Statuswelch <- oneway.test(value ~ Status1, data = data, var.equal = F)
Statuswelch
```

```
##
##  One-way analysis of means (not assuming equal variances)
##
## data:  value and Status1
## F = 98.595, num df = 1.00, denom df = 111.85, p-value < 2.2e-16
```

```
Incomewelch <- oneway.test(value ~ HG_NHG, data = data, var.equal = F)
Incomewelch
```

```
##
##  One-way analysis of means (not assuming equal variances)
##
## data:  value and HG_NHG
## F = 89.419, num df = 1.00, denom df = 140.36, p-value < 2.2e-16
```

```
Mixwelch <- oneway.test(value ~ HG_NHG*Status1, data = data, var.equal = F)
Mixwelch
```

```
##
##  One-way analysis of means (not assuming equal variances)
##
## data:  value and HG_NHG * Status1
## F = 56.567, num df = 3.000, denom df = 58.136, p-value < 2.2e-16
```

```
PostHoc <- aov(value ~ Status1*HG_NHG, data=data)
#Alternativ auch der Name des Zwischenspeichers "mehr" - PostHoc <- aov(mehr)
TukeyHSD(PostHoc)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = value ~ Status1 * HG_NHG, data = data)
##
## $Status1
##           diff       lwr       upr p adj
## NF-F -24.33619 -28.49956 -20.17282     0
##
## $HG_NHG
##               diff       lwr       upr p adj
## Low-High -11.21456 -15.39888 -7.030246 4e-07
##
## $`Status1:HG_NHG`
##                       diff       lwr         upr     p adj
## NF:High-F:High -21.7144985 -28.78225 -14.6467439 0.0000000
## F:Low-F:High   -21.4566038 -32.92025  -9.9929559 0.0000172
## NF:Low-F:High  -32.2330744 -38.75512 -25.7110240 0.0000000
## F:Low-NF:High    0.2578947 -11.55943  12.0752235 0.9999343
## NF:Low-NF:High -10.5185759 -17.64397  -3.3931863 0.0010507
## NF:Low-F:Low   -10.7764706 -22.27574   0.7228009 0.0749616
```

```
library(rstatix)
```

```
##
## Attaching package: 'rstatix'

## The following object is masked from 'package:stats':
##
##     filter
```

```
data['merge'] <- paste(data$Status1,data$HG_NHG)

gm <- rbind(games_howell_test(data, value ~ Status1),games_howell_test(data, value ~ HG_NHG),games_howel
gm
```

```
## # A tibble: 8 x 8
##   .y.   group1  group2  estimate conf.low conf.high    p.adj p.adj.signif
##   <chr> <chr>   <chr>      <dbl>    <dbl>     <dbl>    <dbl> <chr>
## 1 value F       NF       -24.3     -29.2     -19.5  2.35e-14 ****
## 2 value High    Low      -21.4     -25.9     -16.9  4.50e-14 ****
## 3 value F High  F Low    -21.5     -28.0     -14.9  3.77e-11 ****
## 4 value F High  NF High  -21.7     -30.0     -13.4  6.46e- 9 ****
## 5 value F High  NF Low   -32.2     -38.7     -25.8  1.47e-10 ****
## 6 value F Low   NF High   -0.258    -7.31      6.79 1.00e+ 0 ns
## 7 value F Low   NF Low   -10.8     -15.6      -6.00 5.13e- 6 ****
## 8 value NF High NF Low   -10.5     -17.5      -3.51 1.00e- 3 ***
```

```
library(sjstats)
```

```
## Registered S3 methods overwritten by 'lme4':
```

```
##   method                           from
##   cooks.distance.influence.merMod car
##   influence.merMod                 car
##   dfbeta.influence.merMod          car
##   dfbetas.influence.merMod         car

##
## Attaching package: 'sjstats'

## The following object is masked from 'package:psych':
##
##     phi
```

```
eta <- effectsize::eta_squared(multiAnova1, partial = TRUE)
```

```
eta
```

```
## Parameter      | Eta2 (partial) |       90% CI
## -----------------------------------------------
## Status1        |           0.47 | [0.38, 0.55]
## HG_NHG         |           0.19 | [0.10, 0.28]
## Status1:HG_NHG |           0.03 | [0.00, 0.09]
```

```
st <- sqrt(0.47/(1-0.47))

sprintf("effectsize for Freedom Status: f= %.3f", st)
```

```
## [1] "effectsize for Freedom Status: f= 0.942"
```

```
st <- sqrt(0.19/(1-0.19))
sprintf("effectsize for Income Level: f= %.3f", st)
```

```
## [1] "effectsize for Income Level: f= 0.484"
```

```
st <- sqrt(0.03/(1-0.03))
sprintf("effectsize for Income Level X Freedom Status: f= %.3f", st)
```

```
## [1] "effectsize for Income Level X Freedom Status: f= 0.176"
```

According to result there is a relation between Freedom Hause index und Corruption ($F(1, 111.85) = 98.595$, $p = 2.2e-16$). H1 is accepted. Moreover for Income Level there are also differences between Income level Corruption ($F(1,140.36) = 89.419$, $p < 2.2e-16$). H1 is accepted. The Interaction effects of Income Level und Freedom Hause index Corruption is significant ($F(3,58.136) = 56.567$, $p < 2.2e-16$). The effect of Income Leve is dependent on partially Freedom Hause index. H1 for interaction is accepted.

Regarding the Posthoc test, the impact of Freedom and income level is clear, but their interaction should be analyzed carefully. for instance, the differences between non-democratic countries with high-income levels and democratic countries with low-income levels are not significant.

Die effect size ist not only for Freedom Hause index ($f = 0.942$) but also for income Level ($f = 0.484$) according to Cohen (1988) a strong effekt, and for interaction ($f=0.176$) is middle.