Introduction to R

Jaime Undurraga

25 March 2016

## R Markdown

This is an R Markdown presentation. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

## why using R?

* It is free and open source
* Time and processing effitient
* Many libraries
* Continuosly improved and developed by R community
* Well documented (your answer will be usually found in stckoverflow)
* It makes statistical analysis reproducible to others
* It is nice

## R IDE

* R Studio is a nice Interface development enviorment to use R

## Libraries

-R libraris can be installed and loaded

library(xlsx)

## Loading required package: rJava

## Loading required package: xlsxjars

## Data Frames

* Data Frames contain and allow to manipulate data in many ways. Imported data from Excel, SPPS, cvc, txt and other fomars can be loded into R data frames

raw <- read.xlsx('/home/jundurraga/Dropbox/Documents/UCL/Presentations/ARO2016/analysis/questionnaire\_data.xlsx'  
 , 5, encoding = "UTF-8")  
head(raw, n=2)

## subject Age\_years Sex right\_dB\_HL\_250\_Hz right\_dB\_HL\_500\_Hz  
## 1 1WR 48.11781 M 20 20  
## 2 2PP 22.35616 F -5 -5  
## right\_dB\_HL\_1000\_Hz right\_dB\_HL\_2000\_Hz right\_dB\_HL\_4000\_Hz  
## 1 20 15 10  
## 2 -10 -5 10  
## right\_dB\_HL\_8000\_Hz left\_dB\_HL\_250\_Hz left\_dB\_HL\_500\_Hz  
## 1 35 25 50  
## 2 10 -5 -5  
## left\_dB\_HL\_1000\_Hz left\_dB\_HL\_2000\_Hz left\_dB\_HL\_4000\_Hz  
## 1 25 15 20  
## 2 -5 -5 -5  
## left\_dB\_HL\_8000\_Hz right\_ave left\_ave left\_right\_ave better\_ear\_ave  
## 1 25 17 27 22 17  
## 2 5 -3 -5 -4 -5  
## left\_right\_ave\_dB\_HL\_4000\_Hz better\_ear\_dB\_HL\_4000\_Hz  
## 1 15.0 10  
## 2 2.5 -5  
## left\_right\_ave\_dB\_HL\_500\_Hz better\_ear\_dB\_HL\_500\_Hz  
## 1 35 20  
## 2 -5 -5

## Descriptive statistics

summary(raw)

## subject Age\_years Sex right\_dB\_HL\_250\_Hz right\_dB\_HL\_500\_Hz  
## 10AA : 1 Min. :20.09 F:10 Min. :-5.000 Min. :-5.0   
## 11RR : 1 1st Qu.:24.45 M: 9 1st Qu.: 0.000 1st Qu.: 0.0   
## 12SR : 1 Median :35.19 Median : 0.000 Median : 5.0   
## 13BL : 1 Mean :34.43 Mean : 4.211 Mean : 5.0   
## 14PT : 1 3rd Qu.:39.67 3rd Qu.: 5.000 3rd Qu.: 7.5   
## 15SE : 1 Max. :54.00 Max. :30.000 Max. :25.0   
## (Other):13   
## right\_dB\_HL\_1000\_Hz right\_dB\_HL\_2000\_Hz right\_dB\_HL\_4000\_Hz  
## Min. :-10.000 Min. :-5.000 Min. :-5.000   
## 1st Qu.: 0.000 1st Qu.: 0.000 1st Qu.: 0.000   
## Median : 5.000 Median : 0.000 Median : 5.000   
## Mean : 6.053 Mean : 3.158 Mean : 6.316   
## 3rd Qu.: 10.000 3rd Qu.: 5.000 3rd Qu.:10.000   
## Max. : 20.000 Max. :15.000 Max. :30.000   
##   
## right\_dB\_HL\_8000\_Hz left\_dB\_HL\_250\_Hz left\_dB\_HL\_500\_Hz  
## Min. :-5.00 Min. :-5.000 Min. :-10.000   
## 1st Qu.: 2.50 1st Qu.:-5.000 1st Qu.: 0.000   
## Median :10.00 Median : 0.000 Median : 0.000   
## Mean :10.53 Mean : 2.105 Mean : 3.421   
## 3rd Qu.:15.00 3rd Qu.: 5.000 3rd Qu.: 5.000   
## Max. :40.00 Max. :25.000 Max. : 50.000   
##   
## left\_dB\_HL\_1000\_Hz left\_dB\_HL\_2000\_Hz left\_dB\_HL\_4000\_Hz  
## Min. :-5.000 Min. :-5.000 Min. :-10.000   
## 1st Qu.: 0.000 1st Qu.: 0.000 1st Qu.: 0.000   
## Median : 0.000 Median : 0.000 Median : 5.000   
## Mean : 2.632 Mean : 2.895 Mean : 6.842   
## 3rd Qu.: 5.000 3rd Qu.: 5.000 3rd Qu.: 12.500   
## Max. :25.000 Max. :15.000 Max. : 30.000   
##   
## left\_dB\_HL\_8000\_Hz right\_ave left\_ave left\_right\_ave   
## Min. :-10.0 Min. :-3.000 Min. :-5.000 Min. :-4.000   
## 1st Qu.: 2.5 1st Qu.: 1.500 1st Qu.: 0.500 1st Qu.: 1.500   
## Median : 10.0 Median : 3.000 Median : 2.000 Median : 3.000   
## Mean : 10.0 Mean : 4.947 Mean : 3.579 Mean : 4.263   
## 3rd Qu.: 15.0 3rd Qu.: 9.000 3rd Qu.: 5.000 3rd Qu.: 5.750   
## Max. : 40.0 Max. :17.000 Max. :27.000 Max. :22.000   
##   
## better\_ear\_ave left\_right\_ave\_dB\_HL\_4000\_Hz better\_ear\_dB\_HL\_4000\_Hz  
## Min. :-5.000 Min. :-2.500 Min. :-10.000   
## 1st Qu.: 0.500 1st Qu.: 0.000 1st Qu.: -5.000   
## Median : 2.000 Median : 5.000 Median : 5.000   
## Mean : 2.895 Mean : 6.579 Mean : 3.684   
## 3rd Qu.: 4.500 3rd Qu.:11.250 3rd Qu.: 10.000   
## Max. :17.000 Max. :30.000 Max. : 30.000   
##   
## left\_right\_ave\_dB\_HL\_500\_Hz better\_ear\_dB\_HL\_500\_Hz  
## Min. :-5.000 Min. :-10.000   
## 1st Qu.: 0.000 1st Qu.: -2.500   
## Median : 2.500 Median : 0.000   
## Mean : 4.211 Mean : 1.053   
## 3rd Qu.: 5.000 3rd Qu.: 5.000   
## Max. :35.000 Max. : 20.000   
##

## Data frame from wide to long format

library(reshape2)  
raw$subject <- factor(raw$subject)  
PTA <- melt(raw, id.vars=c(1:3,16:23), measure.name=4:16, variable.name = "condition", value.name = "dBHL")  
cnd = read.table(text = as.character(PTA$condition), sep = "\_", colClasses = "character")  
PTA <- cbind(PTA, cnd[c(1,4)])  
names(PTA)[names(PTA) == 'V1'] <- 'EAR'  
names(PTA)[names(PTA) == 'V4'] <- 'Frequency'  
PTA$EAR <- factor(PTA$EAR)  
PTA$Frequency = factor(as.numeric(PTA$Frequency))  
PTA <- PTA[, -4:-12]  
head(PTA, n = 5)

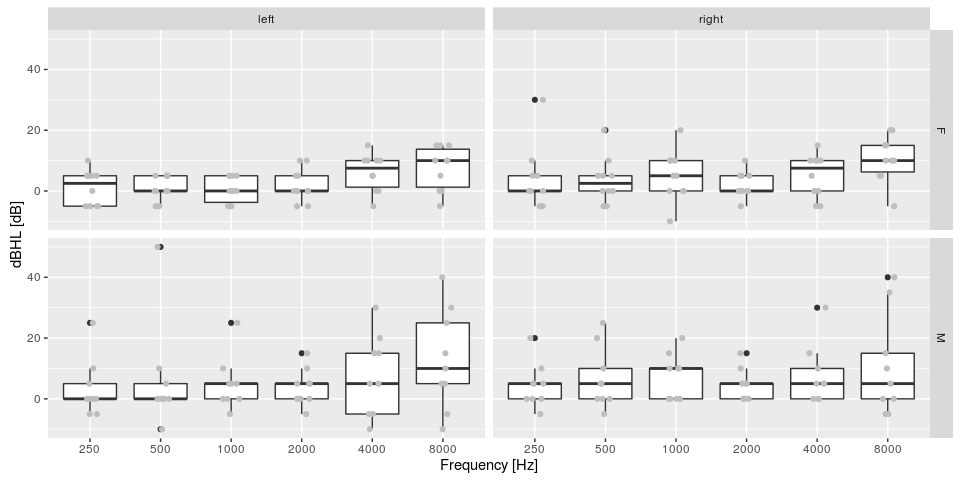
## subject Age\_years Sex dBHL EAR Frequency  
## 1 1WR 48.11781 M 20 right 250  
## 2 2PP 22.35616 F -5 right 250  
## 3 3LS 25.58356 M -5 right 250  
## 4 4BB 39.73973 F 5 right 250  
## 5 5MD 47.80000 M 0 right 250

## Plotting with ggplot

library(ggplot2)  
gp <- (ggplot(data = PTA,  
 aes(x = Frequency,  
 y = dBHL  
 ))  
+ geom\_boxplot(notch=F)  
+ geom\_jitter(mapping=aes(x=Frequency, y=dBHL), width=0.3, height=0, color='gray')  
+ facet\_grid(Sex~EAR)  
+ xlab("Frequency [Hz]")  
+ ylab ("dBHL [dB]"))

## Plotting with ggplot

gp



## Adding new Factors

median\_age <- round(median(raw$Age\_years))  
PTA$AGE\_GROUP <- ifelse(PTA$Age\_years < median\_age, paste("below\_", as.character(median\_age),sep=""),  
 paste("above\_", as.character(median\_age),sep=""))  
PTA$AGE\_GROUP <- factor(PTA$AGE\_GROUP)  
head(PTA[ ,c(1:5, ncol(PTA))], n=10)

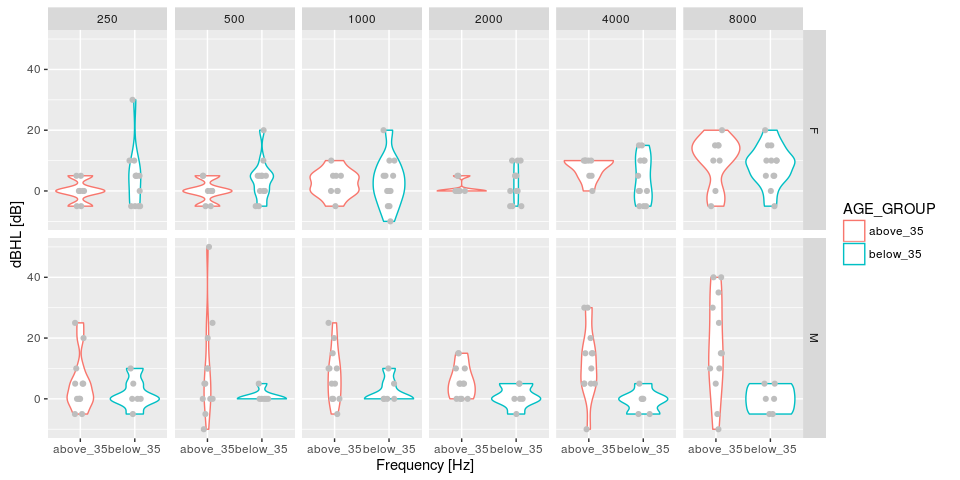
## subject Age\_years Sex dBHL EAR AGE\_GROUP  
## 1 1WR 48.11781 M 20 right above\_35  
## 2 2PP 22.35616 F -5 right below\_35  
## 3 3LS 25.58356 M -5 right below\_35  
## 4 4BB 39.73973 F 5 right above\_35  
## 5 5MD 47.80000 M 0 right above\_35  
## 6 6DD 49.87397 F 0 right above\_35  
## 7 8BV 39.60822 F 0 right above\_35  
## 8 9VF 38.01918 M 5 right above\_35  
## 9 10AA 20.09041 F 10 right below\_35  
## 10 11RR 23.01096 F 30 right below\_35

## Plotting with ggplot

gp <- (ggplot(data = PTA,  
 aes(x = AGE\_GROUP,  
 y = dBHL,  
 color=AGE\_GROUP  
 ))  
+ geom\_violin()  
+ geom\_jitter(mapping=aes(x=AGE\_GROUP, y=dBHL), width=0.3, height=0, color='gray')  
+ facet\_grid(Sex~Frequency)  
+ xlab("Frequency [Hz]")  
+ ylab ("dBHL [dB]"))

## Plotting with ggplot

gp



## ANOVA analysis

library(ez)  
library(pander)  
anv = ezANOVA(data = PTA  
 , dv = .(dBHL)  
 , wid = .(subject)  
 , within = .(Frequency, EAR)  
 , between = .(Sex, AGE\_GROUP)  
 , detailed = T  
 , type = 2  
)

## ANOVA analysis

##   
##   
## | Effect | DFn | DFd | SSn | SSd | F | p | p<.05 | ges |  
## |:----------------------------|:------|:------|:------|:------|:----|:------|:--------|:------|  
## | (Intercept) | 1 | 15 | 6316 | 5280 | 18 | 7e-04 | \* | 0.3 |  
## | Sex | 1 | 15 | 156 | 5280 | 0.4 | 0.5 | | 0.01 |  
## | AGE\_GROUP | 1 | 15 | 669 | 5280 | 2 | 0.2 | | 0.04 |  
## | Frequency | 5 | 75 | 1449 | 5449 | 4 | 0.003 | \* | 0.09 |  
## | EAR | 1 | 15 | 86 | 1245 | 1 | 0.3 | | 0.006 |  
## | Sex:AGE\_GROUP | 1 | 15 | 1185 | 5280 | 3 | 0.09 | | 0.08 |  
## | Sex:Frequency | 5 | 75 | 81 | 5449 | 0.2 | 1 | | 0.006 |  
## | AGE\_GROUP:Frequency | 5 | 75 | 640 | 5449 | 2 | 0.1 | | 0.04 |  
## | Sex:EAR | 1 | 15 | 23 | 1245 | 0.3 | 0.6 | | 0.002 |  
## | AGE\_GROUP:EAR | 1 | 15 | 19 | 1245 | 0.2 | 0.6 | | 0.001 |  
## | Frequency:EAR | 5 | 75 | 97 | 2492 | 0.6 | 0.7 | | 0.007 |  
## | Sex:AGE\_GROUP:Frequency | 5 | 75 | 163 | 5449 | 0.4 | 0.8 | | 0.01 |  
## | Sex:AGE\_GROUP:EAR | 1 | 15 | 33 | 1245 | 0.4 | 0.5 | | 0.002 |  
## | Sex:Frequency:EAR | 5 | 75 | 58 | 2492 | 0.4 | 0.9 | | 0.004 |  
## | AGE\_GROUP:Frequency:EAR | 5 | 75 | 44 | 2492 | 0.3 | 0.9 | | 0.003 |  
## | Sex:AGE\_GROUP:Frequency:EAR | 5 | 75 | 36 | 2492 | 0.2 | 1 | | 0.002 |

## ANOVA Mauchly's Test for Sphericity

##   
##   
## | Effect | W | p | p<.05 |  
## |:----------------------------|:-----|:------|:--------|  
## | Frequency | 0.04 | 9e-05 | \* |  
## | Sex:Frequency | 0.04 | 9e-05 | \* |  
## | AGE\_GROUP:Frequency | 0.04 | 9e-05 | \* |  
## | Sex:AGE\_GROUP:Frequency | 0.04 | 9e-05 | \* |  
## | Frequency:EAR | 0.1 | 0.01 | \* |  
## | Sex:Frequency:EAR | 0.1 | 0.01 | \* |  
## | AGE\_GROUP:Frequency:EAR | 0.1 | 0.01 | \* |  
## | Sex:AGE\_GROUP:Frequency:EAR | 0.1 | 0.01 | \* |

## ANOVA Sphericity Corrections

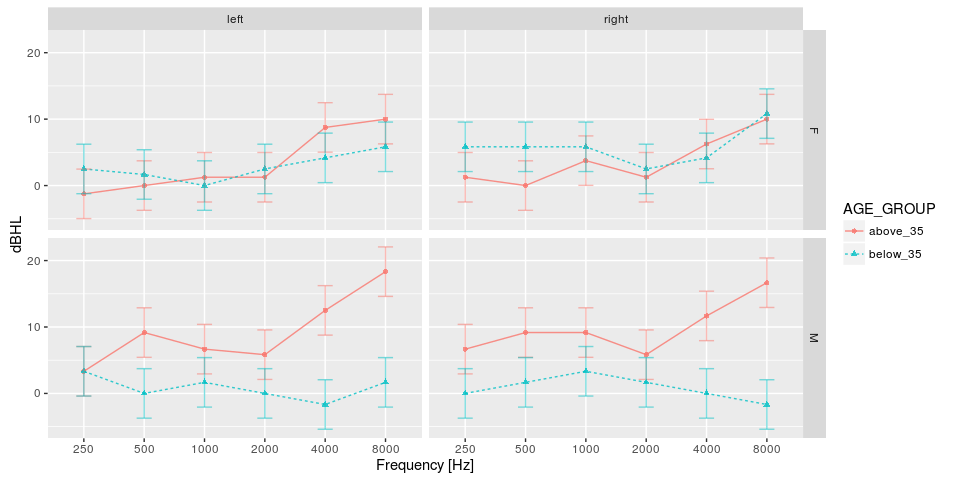
##   
##   
## | Effect | GGe | p[GG] | p[GG]<.05 | HFe | p[HF] | p[HF]<.05 |  
## |:----------------------------|:------|:--------|:------------|:------|:--------|:------------|  
## | Frequency | 0.4 | 0.03 | \* | 0.5 | 0.02 | \* |  
## | Sex:Frequency | 0.4 | 0.8 | | 0.5 | 0.8 | |  
## | AGE\_GROUP:Frequency | 0.4 | 0.2 | | 0.5 | 0.2 | |  
## | Sex:AGE\_GROUP:Frequency | 0.4 | 0.7 | | 0.5 | 0.7 | |  
## | Frequency:EAR | 0.6 | 0.6 | | 0.7 | 0.7 | |  
## | Sex:Frequency:EAR | 0.6 | 0.8 | | 0.7 | 0.8 | |  
## | AGE\_GROUP:Frequency:EAR | 0.6 | 0.8 | | 0.7 | 0.9 | |  
## | Sex:AGE\_GROUP:Frequency:EAR | 0.6 | 0.9 | | 0.7 | 0.9 | |

## ANOVA plots

ez1 <- ezPlot(data = PTA  
 , x = Frequency  
 , dv = .(dBHL)  
 , wid = .(subject)  
 , within = .(Frequency, EAR)  
 , between = .(Sex, AGE\_GROUP)  
 , type = 2  
 , x\_lab = "Frequency [Hz]"  
 , y\_lab = "dBHL"  
 , split = AGE\_GROUP  
 , col = EAR  
 , row = Sex  
 , print\_code = F)

## ANOVA plots

ez1

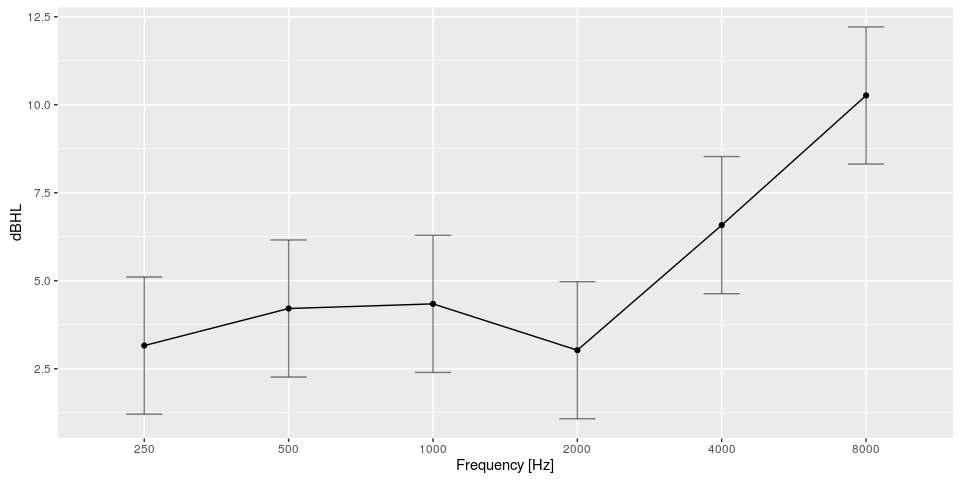


## ANOVA plots

ez2 <- ezPlot(data = PTA  
 , x = Frequency  
 , dv = .(dBHL)  
 , wid = .(subject)  
 , within = .(Frequency)  
 , within\_full = .(Frequency, EAR)  
 , between\_full = .(Sex, AGE\_GROUP)  
 , type = 2  
 , x\_lab = "Frequency [Hz]"  
 , y\_lab = "dBHL"  
 , print\_code = F)

## ANOVA plots

ez2



## Correlations

library(psych)

##   
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':  
##   
## %+%, alpha

library(plyr)  
my\_corr\_f <- function(x,y)  
{  
 r\_corr <- cor.test(x, y)  
 b <- lm(x ~ y)  
 tab\_cor <- data.frame(Value = b$coefficients[2], Std.Error=summary(b)$sigma, t = r\_corr$statistic, df = r\_corr$parameter, p = r\_corr$p.value, r = r\_corr$estimate)  
 row.names(tab\_cor) <- NULL  
 return(tab\_cor)  
}  
corrs <- ddply(PTA, .(Sex, Frequency), function(df) my\_corr\_f(df$Age\_years, df$dBHL))  
pandoc.table(corrs, split.table = Inf, digits = 1, style="rmarkdown", justify = 'left')

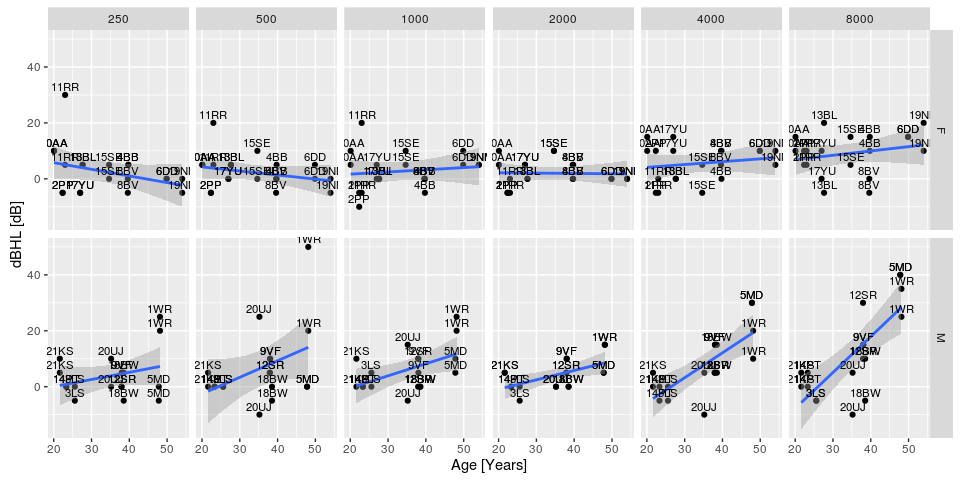
##   
##   
## | Sex | Frequency | Value | Std.Error | t | df | p | r |  
## |:------|:------------|:--------|:------------|:-----|:-----|:------|:------|  
## | F | 250 | -0.5 | 11 | -2 | 18 | 0.2 | -0.3 |  
## | F | 500 | -0.5 | 11 | -1 | 18 | 0.2 | -0.3 |  
## | F | 1000 | 0.2 | 12 | 0.6 | 18 | 0.6 | 0.1 |  
## | F | 2000 | -0.07 | 12 | -0.1 | 18 | 0.9 | -0.03 |  
## | F | 4000 | 0.3 | 12 | 0.8 | 18 | 0.4 | 0.2 |  
## | F | 8000 | 0.4 | 11 | 1 | 18 | 0.3 | 0.2 |  
## | M | 250 | 0.4 | 9 | 1 | 16 | 0.2 | 0.3 |  
## | M | 500 | 0.3 | 9 | 2 | 16 | 0.09 | 0.4 |  
## | M | 1000 | 0.6 | 8 | 2 | 16 | 0.03 | 0.5 |  
## | M | 2000 | 1 | 8 | 3 | 16 | 0.01 | 0.6 |  
## | M | 4000 | 0.6 | 6 | 5 | 16 | 3e-04 | 0.8 |  
## | M | 8000 | 0.4 | 6 | 5 | 16 | 3e-04 | 0.8 |

## Plotting correlations

gp <- (ggplot(data = PTA,  
 aes(x = Age\_years,  
 y = dBHL  
 ))  
+ geom\_point()  
+ geom\_smooth(method="lm", se=T)  
+ geom\_text(aes(label = subject), vjust=-.5, size=3)  
+ facet\_grid(Sex~Frequency)  
+ xlab("Age [Years]")  
+ ylab ("dBHL [dB]"))

## Plotting correlations

gp



## Nonparametric tests

* Independent 2-group Mann-Whitney U Test

mean\_freq <- ddply(PTA, .(subject, Frequency)  
 , summarise  
 , dBHL = mean(dBHL)  
 , AGE\_GROUP = unique(AGE\_GROUP))  
  
wt <- wilcox.test(dBHL ~ AGE\_GROUP, data=mean\_freq)   
tab <- data.frame(W=wt$statistic, p.value=wt$p.value)  
pandoc.table(tab, split.table = Inf, digits = 1, style="rmarkdown", justify = 'left')

##   
##   
## | &nbsp; | W | p.value |  
## |:--------|:-----|:----------|  
## | \*\*W\*\* | 1953 | 0.06 |

* Kruskal Wallis Test One Way Anova by Ranks

kt <- kruskal.test(dBHL ~ AGE\_GROUP, data=mean\_freq)   
tab <- data.frame(chi.squared=kt$statistic, df=kt$parameter ,p.value=kt$p.value)  
pandoc.table(tab, split.table = Inf, digits = 1, style="rmarkdown", justify = 'left')

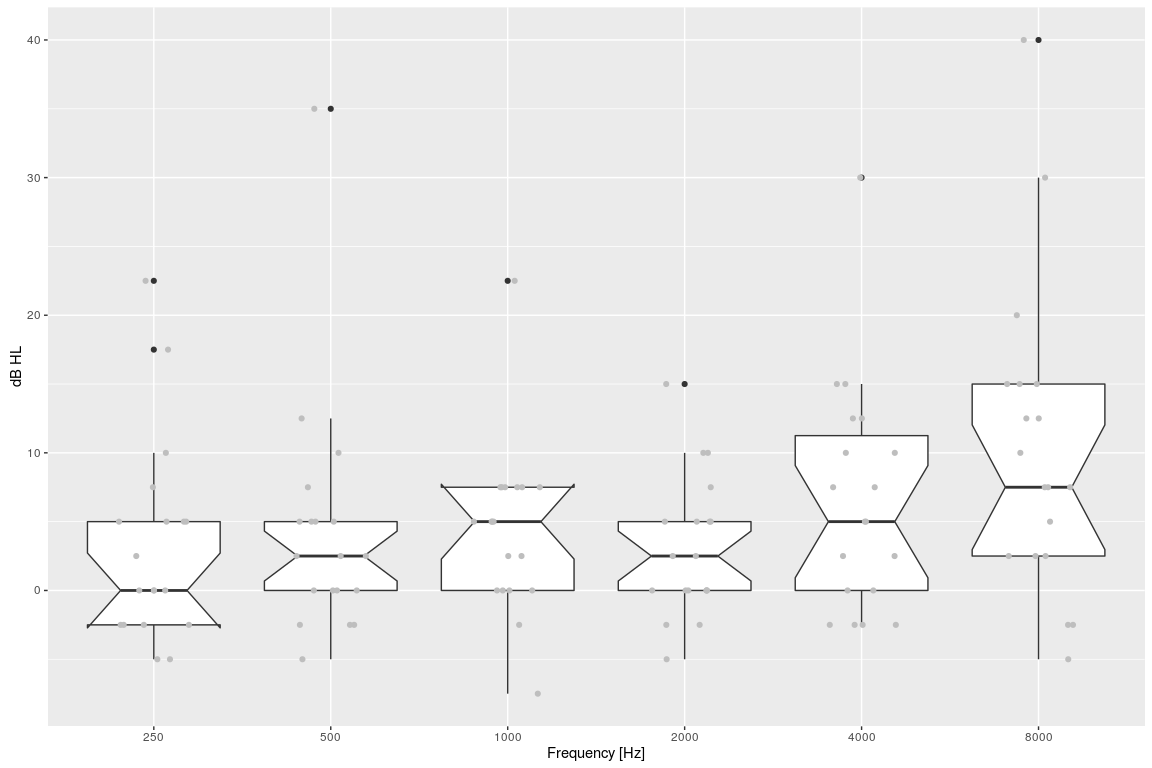
##   
##   
## | &nbsp; | chi.squared | df | p.value |  
## |:---------------------------------|:--------------|:-----|:----------|  
## | \*\*Kruskal-Wallis chi-squared\*\* | 4 | 1 | 0.06 |

kt <- kruskal.test(dBHL ~ Frequency, data=mean\_freq)   
tab <- data.frame(chi.squared=kt$statistic,df=kt$parameter, p.value=kt$p.value)  
pandoc.table(tab, split.table = Inf, digits = 1, style="rmarkdown", justify = 'left')

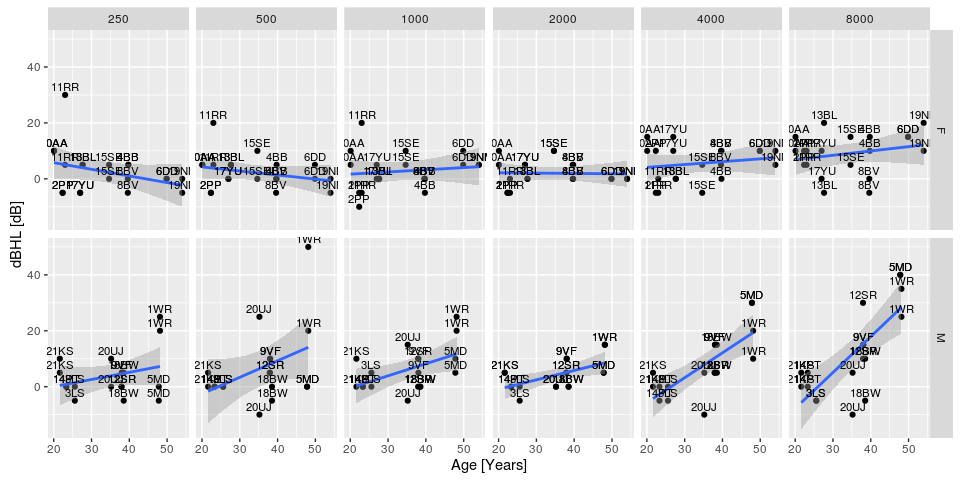
##   
##   
## | &nbsp; | chi.squared | df | p.value |  
## |:---------------------------------|:--------------|:-----|:----------|  
## | \*\*Kruskal-Wallis chi-squared\*\* | 10 | 5 | 0.09 |

(ggplot(data = mean\_freq,  
 aes(x = Frequency,  
 y = dBHL  
 ))  
+ geom\_boxplot(notch=T)  
+ geom\_jitter(mapping=aes(x=Frequency, y=dBHL), width=0.5, height=0, color='gray')  
+ xlab("Frequency [Hz]")  
+ ylab ("dB HL"))

## notch went outside hinges. Try setting notch=FALSE.  
## notch went outside hinges. Try setting notch=FALSE.



gp



## Nonparametric tests

* Friedman Test

ft <- friedman.test(dBHL ~ Frequency | subject, data=mean\_freq)   
tab <- data.frame(chi.squared=ft$statistic,df=ft$parameter, p.value=ft$p.value)  
pandoc.table(tab, split.table = Inf, digits = 1, style="rmarkdown", justify = 'left')

##   
##   
## | &nbsp; | chi.squared | df | p.value |  
## |:---------------------------|:--------------|:-----|:----------|  
## | \*\*Friedman chi-squared\*\* | 12 | 5 | 0.04 |