Exploring the Impact of biopsychosocial Factors on Hearing Loss Incidence Rates: A Population-Based Study.

Keshav Kumar

12 July, 2024

Table of Contents

###### PROLOG ########  
  
  
# PROLOG ###  
# PROJECT: Capstone EHDI Paper   
# PURPOSE: Capstone   
# DIR: OneDrive/Keshav-Personal/Desktop/R\_git/Capstone   
# DATA: ehdi\_acs\_Insur\_data\_2015to2020\_cleaned20240207.csv   
# AUTHOR: Keshav Kumar   
# CREATED: Apr 22, 2024   
# LATEST: JUL 4, 2024   
# NOTES: Source for GOVT spending data:   
# ACS social explorer 2024.  
# Kaiser Family FOundation for healthcare spending data  
  
# PROLOG ###   
  
  
# install packages and open libraries  
  
# stargazer for model comparison  
# sandwich For robust SE estimator  
# broom for getting results with Robust SEs  
# MASS For negative binomial  
# lmtest For model comparison  
# SMPracticals for lung cancer data  
# ggplot2 for plots  
# writing results to excel  
# magrittr for pipes  
# sessioninfo for session\_info at bottom  
# details for session\_info at bottom  
# ggthemes for tufte theme  
# ggrepel for text plotting  
# patchwork for combining plots  
# sjplot for model tables  
pacman::p\_load(stargazer, sandwich, MASS, lmtest, SMPracticals, ggplot2, writexl, broom, broom.mixed, dplyr, tidyverse, lme4, summarytools, tableone, reshape2, kableExtra, magrittr, sessioninfo, details, ggthemes, ggrepel, patchwork, sjPlot, tinytex\_root)  
  
# plot theme  
theme\_set(theme\_tufte()) # but might not carry over in chunks  
  
# Okabe-Ito colorblind-friendly color palette:  
# https://jfly.uni-koeln.de/color/  
  
oi\_pal <- c("#000000", "#E69F00", "#56B4E9", "#009E73", "#F0E442",   
 "#0072B2", "#D55E00", "#CC79A7", "#999999")

### R Markdown

This is an R Markdown document. 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>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

# # Clear existing variables from the environment  
# rm(list = ls ())  
# # Set the working directory to the project folder  
# setwd("C:/Users/kesha/OneDrive/Desktop/R\_git/Capstone")

### Set environment is commented out since this is in already setup in git

# # Set the base folder path  
# box\_folder= "C:/Users/kesha/OneDrive/Desktop/R\_git/Capstone"   
#   
# ## Input directory  
# inputData.dir <- paste(box\_folder,"ACS\_social\_explorer\_data")  
  
## Output directory  
# out.dir <- paste(box\_folder,"/Output/Result", Sys.Date(), sep="")  
# if (!dir.exists(out.dir)) {  
# dir.create(out.dir, recursive = TRUE)  
# }  
# setwd(out.dir)

### Data Management

###### DATA MGMT #####  
  
# Create a dataset for poission  
  
ehdi<- read.csv("ACS\_social\_explorer\_data/ehdi\_acs\_Insur\_data\_2015to2020\_cleaned20240207.csv",header = TRUE)  
  
ehdi\_pos <- ehdi %>%   
 select(`State`,`Year`,`HL`,`Total\_population`, `Black\_under\_5yr`,`White\_under\_5yr`, `Pop\_with\_insur`,`Median\_household\_income`,`Health\_Spending\_per\_Capita`) %>%   
 na.omit() %>%  
 mutate(Year = as.factor(Year)) #%>%   
 # #mutate(PerHL = (HL/Tscr)\*100) %>%   
 # group\_by(Year) %>%  
 # summarise(across(where(is.numeric), ~sum(., na.rm=TRUE))) %>%  
 # mutate(State="Total", Year, sep="\_")  
  
  
#filter states that have data for all the years for all the variables  
  
  
# Check which states have data for all years  
states\_all\_years <- ehdi\_pos %>%  
 group\_by(State) %>%  
 summarise(num\_years = n\_distinct(Year)) %>%  
 filter(num\_years == 6) %>%  
 pull(State)  
  
# Filter the dataset to include only those states  
ehdi\_pos <- ehdi\_pos %>%  
 filter(State %in% states\_all\_years)  
  
#write out filtered data  
#write\_csv(ehdi\_pos,  
# "Output/final\_data/final\_data\_filtered.csv")

### Descriptive Statistics mean ± SD (min-max)

descriptive\_table<- ehdi\_pos %>%  
 group\_by(Year) %>%  
 summarise(  
 `Hearing Loss Cases` = paste(round(mean(HL, na.rm = TRUE), 2), "±", round(sd(HL, na.rm = TRUE), 2), "(", min(HL, na.rm = TRUE), "-", max(HL, na.rm = TRUE), ")", sep = " "),  
 `Black population under 5yr of age` = paste(round(mean(Black\_under\_5yr, na.rm = TRUE) / 1000, 2), "±", round(sd(Black\_under\_5yr, na.rm = TRUE) / 1000, 2), "(", round(min(Black\_under\_5yr, na.rm = TRUE) / 1000, 2), "-", round(max(Black\_under\_5yr, na.rm = TRUE) / 1000, 2), ")", sep = " "),  
 `Population Insured` = paste(round(mean(Pop\_with\_insur, na.rm = TRUE) / 1000, 2), "±", round(sd(Pop\_with\_insur, na.rm = TRUE) / 1000, 2), "(", round(min(Pop\_with\_insur, na.rm = TRUE) / 1000, 2), "-", round(max(Pop\_with\_insur, na.rm = TRUE) / 1000, 2), ")", sep = " "),  
 `Median household income in dollar` = paste(round(mean(Median\_household\_income, na.rm = TRUE) / 1000, 2), "±", round(sd(Median\_household\_income, na.rm = TRUE) / 1000, 2), "(", round(min(Median\_household\_income, na.rm = TRUE) / 1000, 2), "-", round(max(Median\_household\_income, na.rm = TRUE) / 1000, 2), ")", sep = " "),  
 `Health spending per capita in dollar` = paste(round(mean(Health\_Spending\_per\_Capita, na.rm = TRUE) / 1000, 2), "±", round(sd(Health\_Spending\_per\_Capita, na.rm = TRUE) / 1000, 2), "(", round(min(Health\_Spending\_per\_Capita, na.rm = TRUE) / 1000, 2), "-", round(max(Health\_Spending\_per\_Capita, na.rm = TRUE) / 1000, 2), ")", sep = " ")  
 ) %>%   
 ungroup() %>%  
 as.data.frame() # Convert tibble to data.frame  
# Use kable to create the table with a caption  
kable(descriptive\_table, caption = "Descriptive Table", format = "html") %>%  
 kable\_styling(bootstrap\_options = c("striped", "hover"))

Descriptive Table

Year

Hearing Loss Cases

Black population under 5yr of age

Population Insured

Median household income in dollar

Health spending per capita in dollar

2015

133.64 ± 173.48 ( 9 - 1106 )

55.11 ± 67.11 ( 0.22 - 235.5 )

5424.05 ± 6119.41 ( 493.9 - 32338.47 )

54.92 ± 9.04 ( 41.37 - 74.55 )

8.67 ± 1.35 ( 6.11 - 12.5 )

2016

130.49 ± 163.3 ( 4 - 1023 )

54.84 ± 66.95 ( 0.18 - 236.58 )

5540.68 ± 6288.16 ( 500.21 - 33347.8 )

56.31 ± 9.28 ( 42.34 - 76.07 )

8.99 ± 1.39 ( 6.4 - 12.85 )

2017

131.81 ± 162.65 ( 5 - 1011 )

54.51 ± 66.65 ( 0.19 - 236.79 )

5661.47 ± 6470.11 ( 505.49 - 34446.67 )

58.52 ± 9.71 ( 43.81 - 78.92 )

9.29 ± 1.44 ( 6.63 - 13.13 )

2018

129.13 ± 159.3 ( 10 - 962 )

54.28 ± 66.39 ( 0.19 - 235.65 )

5768.45 ± 6622.26 ( 507.04 - 35373.78 )

60.9 ± 10.14 ( 44.92 - 82.6 )

9.6 ± 1.51 ( 7.01 - 13.49 )

2019

121.47 ± 148.3 ( 8 - 915 )

53.75 ± 65.87 ( 0.21 - 235.79 )

5834.16 ± 6711.78 ( 505.67 - 35872.27 )

63.38 ± 10.53 ( 46.71 - 86.42 )

10.01 ± 1.58 ( 7.19 - 13.93 )

2020

126.15 ± 149.01 ( 6 - 901 )

53.03 ± 64.92 ( 0.32 - 237.41 )

5874.56 ± 6755.67 ( 505.71 - 36032.55 )

65.32 ± 10.85 ( 48.04 - 90.84 )

10.56 ± 1.66 ( 7.52 - 14.38 )

Note: The values displayed in the columns for the Black population under 5 years of age, population insured, median household income, and health spending per capita are shown in hundreds for clarity, although the original figures are in thousands.

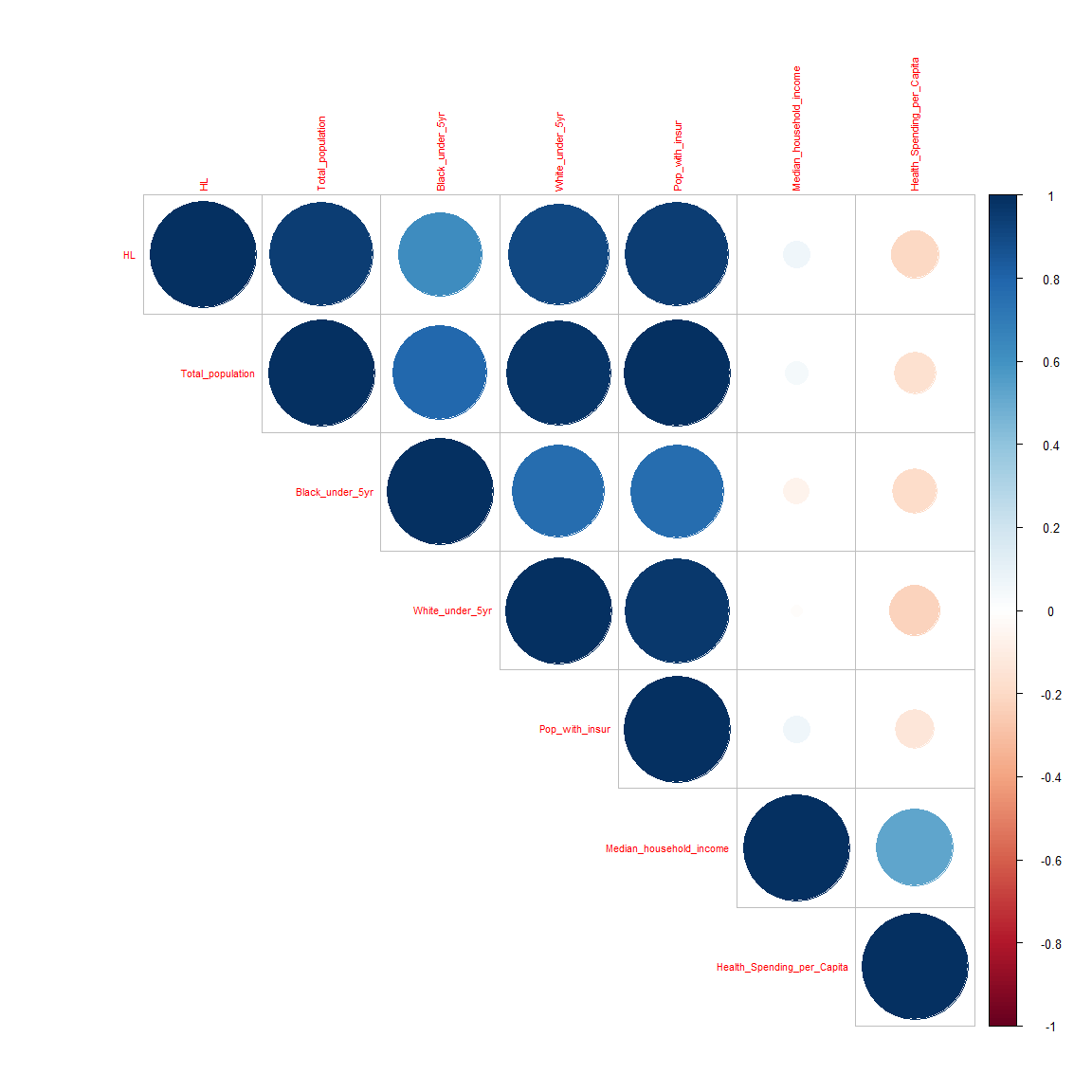
### Correlation

# Get the correlation matrix  
cap\_HL\_ehdi <- ehdi\_pos %>%   
 select(-c(State, Year))  
# Check correlation matrix  
cor\_matrix <- cor(cap\_HL\_ehdi)  
summary(cor\_matrix)

HL Total\_population Black\_under\_5yr White\_under\_5yr

Min. :-0.2025 Min. :-0.1626 Min. :-0.1816 Min. :-0.2289  
1st Qu.: 0.3462 1st Qu.: 0.4159 1st Qu.: 0.2827 1st Qu.: 0.3753  
Median : 0.9027 Median : 0.9456 Median : 0.7630 Median : 0.9027  
Mean : 0.6124 Mean : 0.6552 Mean : 0.5283 Mean : 0.6229  
3rd Qu.: 0.9471 3rd Qu.: 0.9857 3rd Qu.: 0.7758 3rd Qu.: 0.9677  
Max. : 1.0000 Max. : 1.0000 Max. : 1.0000 Max. : 1.0000  
Pop\_with\_insur Median\_household\_income Health\_Spending\_per\_Capita Min. :-0.1390 Min. :-0.06098 Min. :-0.22886  
1st Qu.: 0.4184 1st Qu.: 0.01850 1st Qu.:-0.19205  
Median : 0.9486 Median : 0.06595 Median :-0.16262  
Mean : 0.6580 Mean : 0.23379 Mean : 0.08746  
3rd Qu.: 0.9798 3rd Qu.: 0.29728 3rd Qu.: 0.19389  
Max. : 1.0000 Max. : 1.00000 Max. : 1.00000

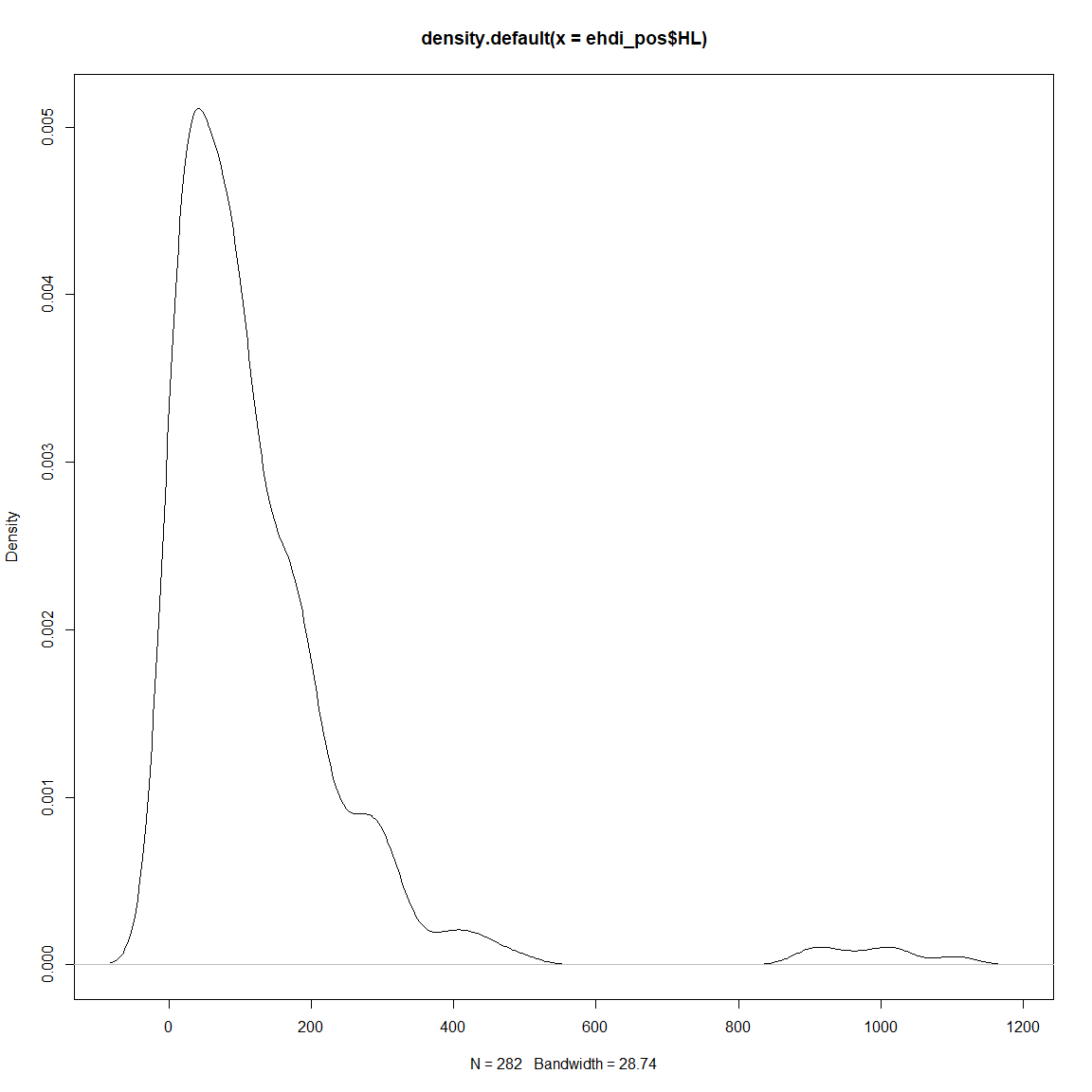
# Visualize correlation matrix  
cor\_plot\_HL <- corrplot::corrplot(cor\_matrix, method = "circle", type = "upper", tl.cex = 0.7)



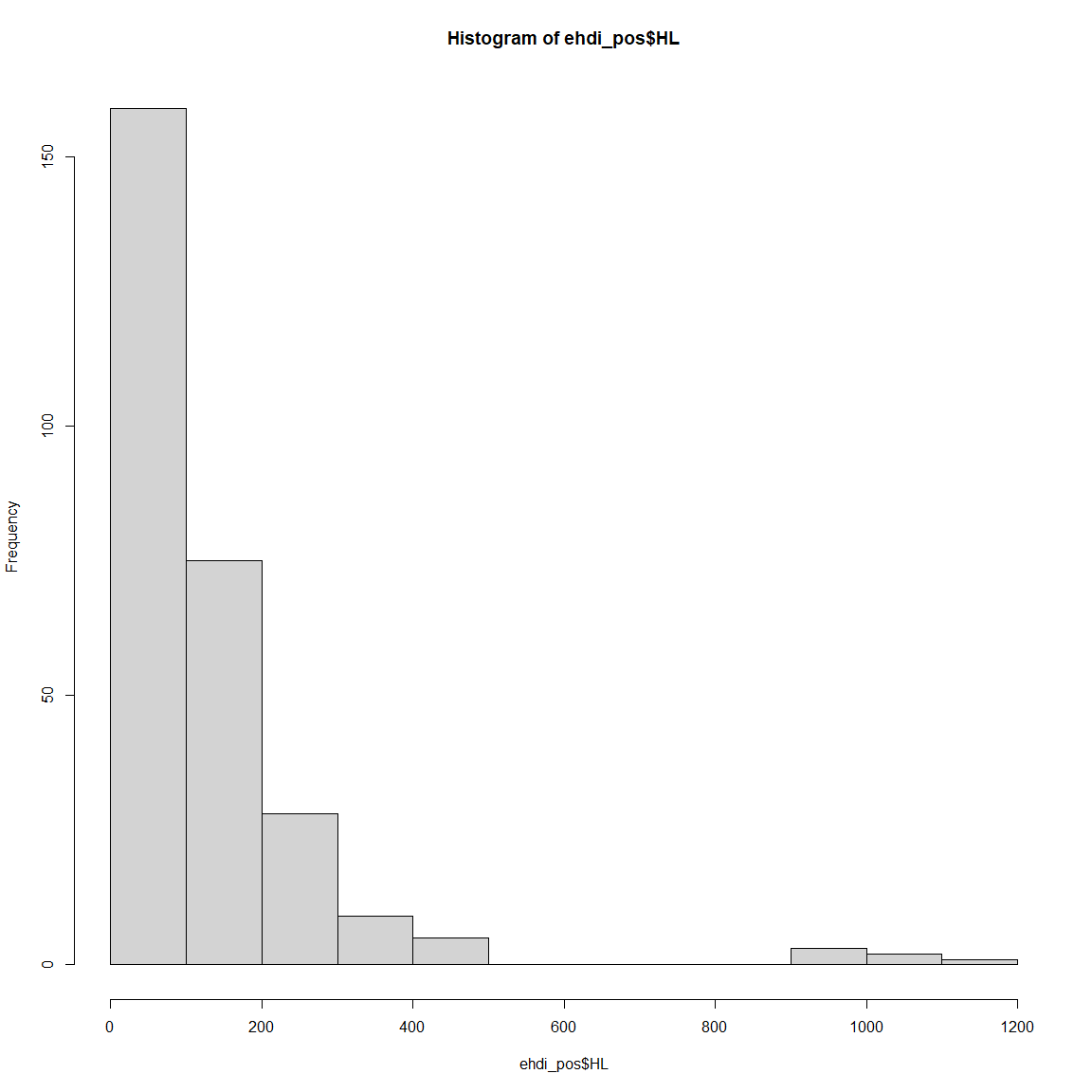
# continuing the rest of the code -----------------------  
# Plot heatmap  
correlation\_plot<- ggplot(data = melt(cor\_matrix), aes(Var1, Var2, fill = value)) +  
 geom\_tile(color = "white") +  
 scale\_fill\_gradient2(low = "blue", high = "red", mid = "white",   
 midpoint = 0, limit = c(-1, 1), space = "Lab",  
 name="Correlation") +  
 theme\_minimal() +  
 theme(axis.text.x = element\_text(angle = 45, vjust = 1,   
 size = 10, hjust = 1))  
#save the plot  
# ggsave(correlation\_plot,  
# # file path  
# file="Output/correlation\_plot\_HL\_capstone.png",  
# # figure size  
# width=1200/96,  
# height=900/96,  
# bg = "white")

### Check Outcome Variable Distribution

# Check shape of distribution of counts of cases using density plot and histogram  
d <- density(ehdi\_pos$HL)   
plot(d)



hist(ehdi\_pos$HL)



#with ggplot  
gghisto <- ggplot(ehdi\_pos, aes(HL))+  
 geom\_histogram(color="black", fill="#3182bd")  
  
ggdens <- ggplot(ehdi\_pos, aes(HL))+  
 geom\_density()  
  
# #save the plot  
# ggsave(gghisto,  
# # file path  
# file="Output/gghisto\_HL\_capstone.png",  
# # figure size  
# width=1200/96,  
# height=900/96,  
# bg = "white")  
#   
# ggsave(ggdens,  
# # file path  
# file="Output/ggdens\_HL\_capstone.png",  
# # figure size  
# width=1200/96,  
# height=900/96,  
# bg = "white")

Our study aims to examine the temporal trends in hearing loss patterns over a six-year period, from 2015 to 2020, taking into account key covariates such as the population of Black infants under 5 years of age, insurance coverage, median household income, and state-level health spending per capita. Additionally, this study seeks to identify any potential shifts in hearing loss patterns attributable to the COVID-19 pandemic.

However, the population sizes (n) differ between the states so we want to account for that by pedicting the log(count/n) from our model. When we exponentiate the beta for state, we will then get the incidence rate ratio. To account for differences in n, we need to use an \*\*offset\*, which is defined by log(n) in the model.

### Create an Offset term

# Add an offset term  
ehdi\_pos\_3 <- ehdi\_pos %>%  
 mutate(offset = log(Total\_population))  
  
ehdi\_pos\_3$State <- factor(ehdi\_pos\_3$State) # Convert State to factor  
contrasts(ehdi\_pos\_3$Year) <- contr.treatment(levels(ehdi\_pos\_3$Year))

### Convert Variables State & Year to the appropriate formats

#converting state to factor  
ehdi\_pos\_3$State <- factor(ehdi\_pos\_3$State)   
  
# convert year to categorical and setting contrasts using treatment coding to compare each year against the baseline category  
contrasts(ehdi\_pos\_3$Year) <- contr.treatment(levels(ehdi\_pos\_3$Year))

The scale() function in R standardizes data by centering and scaling: it first subtracts the mean (centering) and then divides by the standard deviation (scaling) of each variable. Mathematically, for each element x in a variable, the standardized value z is calculated as z = (x−μ)/σ where μ is the mean and σ is the standard deviation of the variable. This process transforms the data to have a mean of 0 and a standard deviation of 1, facilitating comparisons and analyses across different scales.

### Standardization

# Standardize the predictor variables  
ehdi\_pos\_3$Black\_under\_5yr\_std <- scale(ehdi\_pos\_3$Black\_under\_5yr, center = TRUE, scale = TRUE)  
ehdi\_pos\_3$Median\_household\_income\_std <- scale(ehdi\_pos\_3$Median\_household\_income, center = TRUE, scale = TRUE)  
ehdi\_pos\_3$Pop\_with\_insur\_std <- scale(ehdi\_pos\_3$Pop\_with\_insur, center = TRUE, scale = TRUE)  
ehdi\_pos\_3$Health\_Spending\_per\_Capita <- scale(ehdi\_pos\_3$Health\_Spending\_per\_Capita, center = TRUE, scale = TRUE)

### Possion Model

# Fit the Poisson regression model with standardized predictor variables  
HL\_model <- glmer(HL ~ Year + Black\_under\_5yr\_std + Median\_household\_income\_std +   
 Pop\_with\_insur\_std + Health\_Spending\_per\_Capita + (1 | State),   
 data = ehdi\_pos\_3, family = poisson, offset = offset)  
  
# Display the summary of the model  
tab\_model(HL\_model)

HL

Predictors

Incidence Rate Ratios

CI

p

(Intercept)

0.00

0.00 – 0.00

<0.001

Year [2016]

1.00

0.96 – 1.04

0.890

Year [2017]

1.03

0.98 – 1.08

0.217

Year [2018]

1.04

0.97 – 1.11

0.254

Year [2019]

1.01

0.93 – 1.10

0.853

Year [2020]

1.08

0.98 – 1.20

0.137

Black under 5yr std

0.94

0.84 – 1.06

0.297

Median household incomestd

0.95

0.86 – 1.04

0.258

Pop with insur std

0.93

0.83 – 1.03

0.147

Health Spending perCapita

0.92

0.85 – 0.99

0.027

Random Effects

σ2

10.81

τ00 State

0.10

ICC

0.01

N State

47

Observations

282

Marginal R2 / Conditional R2

0.002 / 0.011

### IRRs and 95% CIs

df<- tidy(HL\_model)  
df$IRR<- exp(df$estimate)  
df$lowCI <- exp(df$estimate - 1.96\*df$std.error)  
df$highCI <- exp(df$estimate + 1.96\*df$std.error)  
knitr::kable(df, caption = "Incidence Rate Ratios and 95% Confidence Intervals for Possion")

Incidence Rate Ratios and 95% Confidence Intervals for Possion

effect

group

term

estimate

std.error

statistic

p.value

IRR

lowCI

highCI

fixed

NA

(Intercept)

-10.8297172

0.0532918

-203.2155907

0.0000000

0.0000198

0.0000178

0.0000220

fixed

NA

Year2016

-0.0026921

0.0194337

-0.1385294

0.8898220

0.9973115

0.9600382

1.0360319

fixed

NA

Year2017

0.0299556

0.0242692

1.2343039

0.2170897

1.0304087

0.9825421

1.0806073

fixed

NA

Year2018

0.0369193

0.0323986

1.1395330

0.2544809

1.0376093

0.9737682

1.1056358

fixed

NA

Year2019

0.0079117

0.0427121

0.1852326

0.8530466

1.0079430

0.9269979

1.0959563

fixed

NA

Year2020

0.0790497

0.0531303

1.4878466

0.1367913

1.0822581

0.9752264

1.2010366

fixed

NA

Black\_under\_5yr\_std

-0.0612076

0.0587209

-1.0423480

0.2972504

0.9406279

0.8383660

1.0553636

fixed

NA

Median\_household\_income\_std

-0.0532457

0.0470930

-1.1306516

0.2582017

0.9481470

0.8645484

1.0398292

fixed

NA

Pop\_with\_insur\_std

-0.0770371

0.0531471

-1.4495074

0.1471959

0.9258555

0.8342641

1.0275026

fixed

NA

Health\_Spending\_per\_Capita

-0.0830769

0.0376395

-2.2071724

0.0273020

0.9202804

0.8548320

0.9907397

ran\_pars

State

sd\_\_(Intercept)

0.3116969

NA

NA

NA

1.3657407

NA

NA

### IRR table

# Combine IRR, lowCI, and highCI into a formatted string  
df$`IRR (Low CI - High CI)` <- paste0(round(df$IRR, 2), " (", round(df$lowCI, 2), "-", round(df$highCI, 2), ")")  
df$`p-value` <- round(df$p.value, 2)  
# Select only the formatted IRR and p.value columns  
df\_final <- df[, c("term", "IRR (Low CI - High CI)", "p-value")]  
  
# Print the final adjusted data frame  
knitr::kable(df\_final, caption = "IRR table for Possion")

IRR table for Possion

term

IRR (Low CI - High CI)

p-value

(Intercept)

0 (0-0)

0.00

Year2016

1 (0.96-1.04)

0.89

Year2017

1.03 (0.98-1.08)

0.22

Year2018

1.04 (0.97-1.11)

0.25

Year2019

1.01 (0.93-1.1)

0.85

Year2020

1.08 (0.98-1.2)

0.14

Black\_under\_5yr\_std

0.94 (0.84-1.06)

0.30

Median\_household\_income\_std

0.95 (0.86-1.04)

0.26

Pop\_with\_insur\_std

0.93 (0.83-1.03)

0.15

Health\_Spending\_per\_Capita

0.92 (0.85-0.99)

0.03

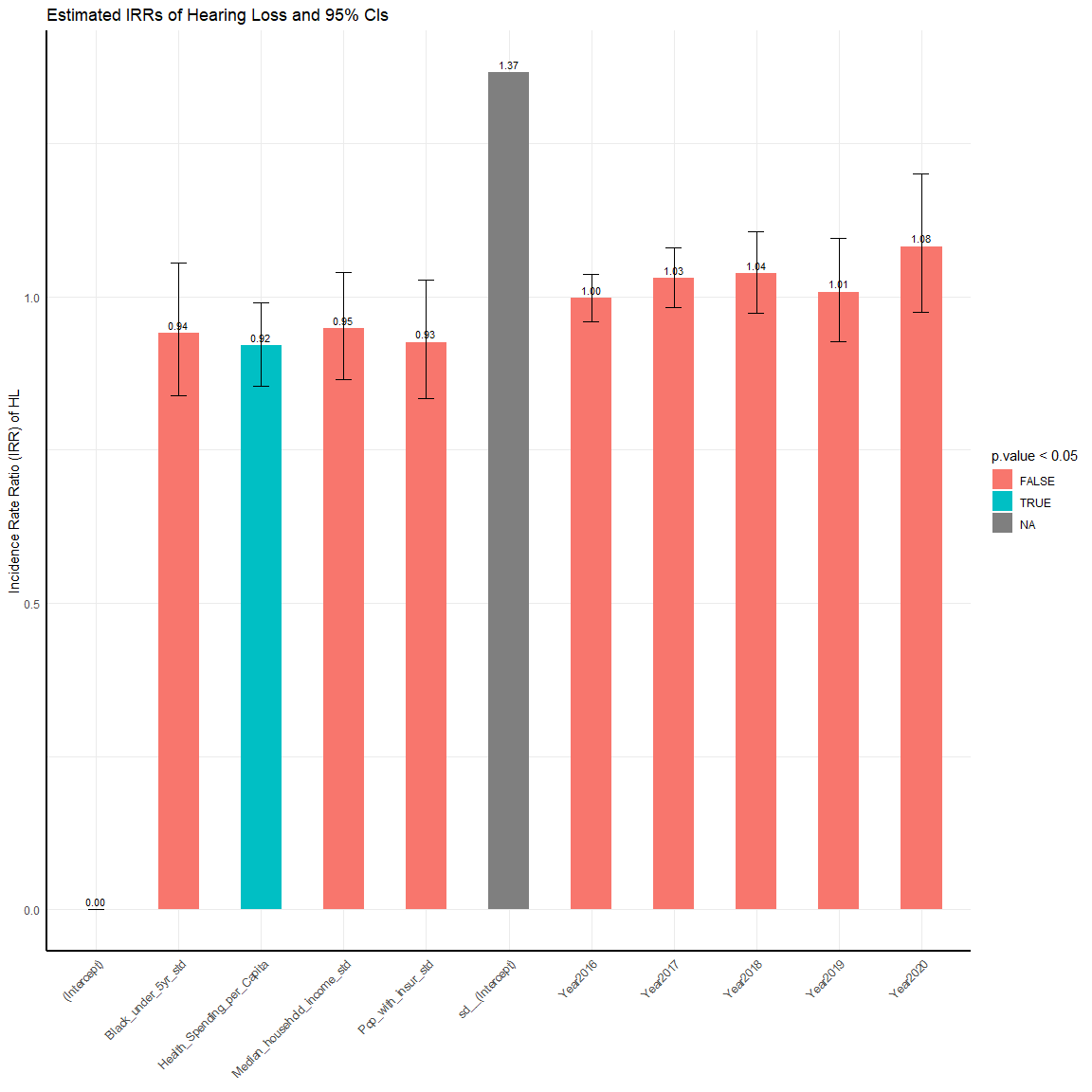
sd\_\_(Intercept)

1.37 (NA-NA)

NA

### Vizualize the Possion Results

plot\_HL<- ggplot(df, aes(x = term, y = IRR, fill = p.value < 0.05, ymin = lowCI, ymax = highCI)) +  
 geom\_bar(stat = "identity", width = 0.5) +  
 geom\_errorbar(width = 0.2, position = position\_dodge(0.5)) +  
 geom\_text(aes(label = sprintf("%.2f", IRR)), vjust = -0.5, size = 3, position = position\_dodge(0.5)) +  
 labs(x = NULL, y = "Incidence Rate Ratio (IRR) of HL", title = "Estimated IRRs of Hearing Loss and 95% CIs") +  
 #scale\_fill\_manual(values = c("FALSE" = "skyblue", "TRUE" = "red"), guide = FALSE) +  
 theme\_minimal() +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1),  
 axis.line = element\_line(size = 1))  
  
print(plot\_HL)



# ggsave(plot\_HL,  
# # file path  
# file="Output/ggplot\_HL\_capstone.png",  
# # figure size  
# width=1200/96,  
# height=900/96,  
# bg = "white")

We can test for overdispersion by running a negative binomial model and then running the LR test. Note for the glm.nb function, there is no offset option so you need to add the offset as a term in the model. Poisson accepts either as above but the glm.nb function does not.

### Negative Binomial

# Fit the negative binomial regression model with standardized predictor variables  
HL\_model\_nb <- glmer.nb(HL ~ Year + Black\_under\_5yr\_std + Median\_household\_income\_std +   
 Pop\_with\_insur\_std + Health\_Spending\_per\_Capita + (1 | State),   
 data = ehdi\_pos\_3, offset = offset)  
  
# Display the summary of the model  
tab\_model(HL\_model\_nb)

HL

Predictors

Incidence Rate Ratios

CI

p

(Intercept)

0.00

0.00 – 0.00

<0.001

Year [2016]

0.98

0.93 – 1.05

0.624

Year [2017]

1.00

0.93 – 1.07

0.990

Year [2018]

0.98

0.91 – 1.06

0.621

Year [2019]

0.94

0.86 – 1.04

0.236

Year [2020]

1.01

0.90 – 1.13

0.872

Black under 5yr std

0.88

0.79 – 0.99

0.029

Median household incomestd

1.07

0.98 – 1.16

0.135

Pop with insur std

1.02

0.91 – 1.14

0.746

Health Spending perCapita

0.86

0.79 – 0.93

<0.001

Random Effects

σ2

10.80

τ00 State

0.07

ICC

0.01

N State

47

Observations

282

Marginal R2 / Conditional R2

0.002 / 0.008

### IRR and CI

df1<- tidy(HL\_model\_nb)  
df1$IRR<- exp(df1$estimate)  
df1$lowCI <- exp(df1$estimate - 1.96\*df1$std.error)  
df1$highCI <- exp(df1$estimate + 1.96\*df1$std.error)  
  
knitr:: kable(df1, caption = "Incidence Rate Ratios and 95% Confidence Intervals for Negative Binomial")

Incidence Rate Ratios and 95% Confidence Intervals for Negative Binomial

effect

group

term

estimate

std.error

statistic

p.value

IRR

lowCI

highCI

fixed

NA

(Intercept)

-10.7911694

0.0494731

-218.1219133

0.0000000

0.0000206

0.0000187

0.0000227

fixed

NA

Year2016

-0.0152940

0.0311931

-0.4903011

0.6239208

0.9848223

0.9264154

1.0469116

fixed

NA

Year2017

-0.0004112

0.0344812

-0.0119239

0.9904863

0.9995889

0.9342657

1.0694795

fixed

NA

Year2018

-0.0198870

0.0402726

-0.4938102

0.6214402

0.9803094

0.9059047

1.0608253

fixed

NA

Year2019

-0.0577906

0.0487745

-1.1848536

0.2360753

0.9438475

0.8577963

1.0385312

fixed

NA

Year2020

0.0094356

0.0586197

0.1609632

0.8721224

1.0094803

0.8999114

1.1323897

fixed

NA

Black\_under\_5yr\_std

-0.1266258

0.0579193

-2.1862445

0.0287977

0.8810633

0.7865117

0.9869816

fixed

NA

Median\_household\_income\_std

0.0657388

0.0439792

1.4947680

0.1349750

1.0679477

0.9797473

1.1640882

fixed

NA

Pop\_with\_insur\_std

0.0185130

0.0572244

0.3235160

0.7463045

1.0186854

0.9106043

1.1395949

fixed

NA

Health\_Spending\_per\_Capita

-0.1500200

0.0421063

-3.5628906

0.0003668

0.8606908

0.7925115

0.9347355

ran\_pars

State

sd\_\_(Intercept)

0.2553348

NA

NA

NA

1.2908938

NA

NA

# write.csv(df1,  
# # file path  
# file="Output/HL\_model\_nb\_IRR\_tbl.csv")

calculating the SD for the significant predictors before standardization for interpretation purpose

### SD for Predictors

# Calculate SD for the black population under 5 years  
sd\_black\_population <- sd(ehdi\_pos$Black\_under\_5yr, na.rm = TRUE)  
  
# Calculate SD for health spending per capita  
sd\_health\_spending <- sd(ehdi\_pos$Health\_Spending\_per\_Capita, na.rm = TRUE)  
  
# Print the standard deviations  
knitr::kable(sd\_black\_population)

x

65731.74

knitr::kable(sd\_health\_spending)

x

1606.917

### IRR Table

# Combine IRR, lowCI, and highCI into a formatted string  
df1$`IRR (Low CI - High CI)` <- paste0(round(df1$IRR, 2), " (", round(df1$lowCI, 2), "-", round(df1$highCI, 2), ")")  
df1$`p-value` <- round(df1$p.value, 2)  
# Select only the formatted IRR and p.value columns  
df1\_final <- df1[, c("term", "IRR (Low CI - High CI)", "p-value")]  
  
# Print the final adjusted dataframe  
knitr::kable(df1\_final, caption = "IRR table for Negative Binomial")

IRR table for Negative Binomial

term

IRR (Low CI - High CI)

p-value

(Intercept)

0 (0-0)

0.00

Year2016

0.98 (0.93-1.05)

0.62

Year2017

1 (0.93-1.07)

0.99

Year2018

0.98 (0.91-1.06)

0.62

Year2019

0.94 (0.86-1.04)

0.24

Year2020

1.01 (0.9-1.13)

0.87

Black\_under\_5yr\_std

0.88 (0.79-0.99)

0.03

Median\_household\_income\_std

1.07 (0.98-1.16)

0.13

Pop\_with\_insur\_std

1.02 (0.91-1.14)

0.75

Health\_Spending\_per\_Capita

0.86 (0.79-0.93)

0.00

sd\_\_(Intercept)

1.29 (NA-NA)

NA

# write.csv(df1\_final,  
# # file path  
# file="Output/neg\_binomial\_IRR\_table.csv")

### Likelihood ratio test

#run lrtest to compare models  
kable(anova(HL\_model, HL\_model\_nb))# if p-value is <0.05 use negbin

npar

AIC

BIC

logLik

deviance

Chisq

Df

Pr(>Chisq)

HL\_model

11

2459.969

2500.030

-1218.984

2437.969

NA

NA

NA

HL\_model\_nb

12

2387.642

2431.345

-1181.821

2363.642

74.32684

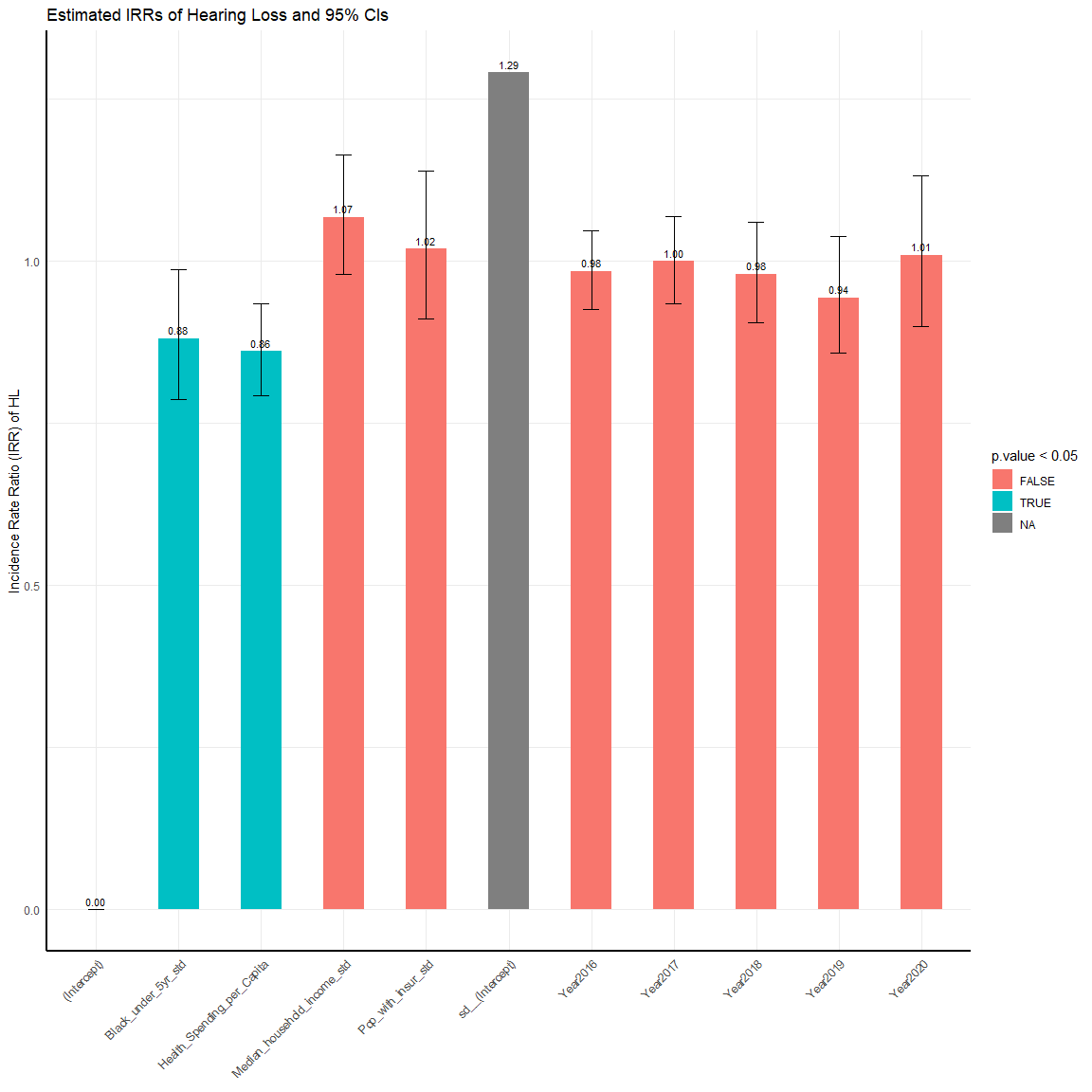
1

0

The LR test indicates that the negative binomial model significantly improve the fit to the data and that we have violated the mean = variance assumption. Therefore the neg binomial model adequately fits these data.

### Vizualize Negative Binomial

plot\_HL\_nb<- ggplot(df1, aes(x = term, y = IRR, fill = p.value < 0.05, ymin = lowCI, ymax = highCI)) +  
 geom\_bar(stat = "identity", width = 0.5) +  
 geom\_errorbar(width = 0.2, position = position\_dodge(0.5)) +  
 geom\_text(aes(label = sprintf("%.2f", IRR)), vjust = -0.5, size = 3, position = position\_dodge(0.5)) +  
 labs(x = NULL, y = "Incidence Rate Ratio (IRR) of HL", title = "Estimated IRRs of Hearing Loss and 95% CIs") +  
 #scale\_fill\_manual(values = c("FALSE" = "skyblue", "TRUE" = "red"), guide = FALSE) +  
 theme\_minimal() +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1),  
 axis.line = element\_line(size = 1))  
  
print(plot\_HL\_nb)



ggsave(plot\_HL\_nb,  
 # file path  
 file="Output/ggplot\_HL\_nb\_capstone.png",  
 # figure size  
 width=1200/96,  
 height=900/96,  
 bg = "white")

Below we further compare the estimates between the two models. The Poisson regression estimates SEs that are usually smaller than those from the negbin.

### Compare Model

model <- stargazer(HL\_model, HL\_model\_nb, title="Model Comparison",  
 type="text", align=TRUE, single.row=TRUE, digits=6)  
  
print(model, caption= "Model Comparison")