missile\_test\_code.R

Tom Brailey

2019-08-03

### Missile Test Temporal Study ###  
 ### POLI 170A ###  
  
rm(list = ls())  
  
# Setup  
setwd(paste0(here::here(), '/data/'))  
wd <- setwd(paste0(here::here(), '/data/'))  
  
  
# Install required packages  
library(magrittr)  
library(ggplot2)  
  
# Install NK xlsx doc  
files <- list.files(wd, "north\_korea\_missile\_test\_database.xlsx")  
files <- files[]  
  
read\_excel\_allsheets <- function(filename) {   
 sheets <- readxl::excel\_sheets(filename)   
 x <- lapply(sheets, function(X) readxl::read\_excel(filename, sheet = X))   
 names(x) <- sheets   
 x   
}   
  
out <- lapply(files, read\_excel\_allsheets)

## New names:  
## \* `` -> ...3  
## \* `` -> ...4  
## \* `` -> ...5  
## New names:  
## \* `` -> ...3  
## \* `` -> ...4  
## \* `` -> ...5

basename(files)

## [1] "north\_korea\_missile\_test\_database.xlsx"

nk\_missile\_tests <- out[[1]]$`Missile Tests`  
nk\_data\_summary <- out[[1]]$`Data Summary`  
nk\_missile\_summary <- out[[1]]$`Missile Summary`  
nk\_facilities <- out[[1]]$`Facilities`  
  
# Install IRN xlsx doc  
files <- list.files(wd, "iran\_missile\_launch\_database.xlsx")  
files <- files[]  
  
out <- lapply(files, read\_excel\_allsheets)

## New names:  
## \* `` -> ...3  
## \* `` -> ...4  
## \* `` -> ...5  
## New names:  
## \* `` -> ...3  
## \* `` -> ...4  
## \* `` -> ...5

basename(files)

## [1] "iran\_missile\_launch\_database.xlsx"

irn\_missile\_tests <- out[[1]]$`Iran Database`  
irn\_data\_summary <- out[[1]]$`Data Summary`  
irn\_missile\_summary <- out[[1]]$`Missile Summary`  
  
rm(out, files, read\_excel\_allsheets)  
  
# Install IRQ data  
irq\_missile\_tests <- rio::import("iraq\_missile\_launch\_database.csv")  
colnames(irq\_missile\_tests) <- as.character(irq\_missile\_tests[1,])  
irq\_missile\_tests <- irq\_missile\_tests[-1,]  
  
# Install PAK data  
pak\_missile\_tests <- rio::import("pakistan\_missile\_launch\_database.csv")  
  
# Clean data  
nk\_missile\_tests$Country <- "North Korea"  
nk\_missile\_tests <- nk\_missile\_tests %>%  
 dplyr::rename(EventID = F1,  
 MissileFamily = `Missile Type`,  
 FacilityLatitude = `Facility Latitude`,  
 Confirmation = `Confirmation Status`,  
 FacilityLongitude = `Facility Longitude`,  
 TestOutcome = `Test Outcome`,  
 DateEntered = `Date Entered/Updated`,   
 FacilityName = `Facility Name`,  
 LandingLocation = `Landing Location`,  
 Source = `Source(s)`,  
 MissileName = `Missile Name`,  
 DistanceTravelled = `Distance Travelled`,  
 AdditionalInformation = `Additional Information`,  
 OtherName = `Other Name`,  
 FacilityLocation = `Facility Location`,  
 LaunchAgency = `Launch Agency/Authority`,  
 LaunchTimeUTC = `Launch Time (UTC)`)  
nk\_missile\_tests$Confirmation[nk\_missile\_tests$Confirmation == "Confirmed"] <- TRUE  
nk\_missile\_tests$Confirmation[nk\_missile\_tests$Confirmation == "Unconfirmed"] <- FALSE  
nk\_missile\_tests$Confirmation <- as.logical(nk\_missile\_tests$Confirmation)  
nk\_missile\_tests$Apogee <- gsub("[a-zA-Z/, ]", "", nk\_missile\_tests$Apogee)  
nk\_missile\_tests$Apogee <- as.numeric(nk\_missile\_tests$Apogee)  
nk\_missile\_tests$DistanceTravelled <- gsub("[a-zA-Z/, ]", "", nk\_missile\_tests$DistanceTravelled)  
nk\_missile\_tests$DistanceTravelled <- as.numeric(nk\_missile\_tests$DistanceTravelled)  
nk\_missile\_tests$Date <- as.character(nk\_missile\_tests$Date)  
  
irn\_missile\_tests$Country <- "Iran"  
irn\_missile\_tests <- irn\_missile\_tests %>%   
 dplyr::rename(Date = DateOccurred)  
irn\_missile\_tests$Date <- as.character(irn\_missile\_tests$Date)  
  
irq\_missile\_tests$Country <- "Iraq"  
irq\_missile\_tests <- irq\_missile\_tests %>%  
 dplyr::rename(MissileName = "",  
 AdditionalInformation = Status,  
 MaxRange = `Maximum Range (km)`,  
 PayloadKG = `Payload (kg)`)  
irq\_missile\_tests$Date[irq\_missile\_tests$Date == "05-Dec-89"] <- "1989-10-05"  
irq\_missile\_tests$Date[irq\_missile\_tests$Date == "Jun-00"] <- "2000-06-01"  
irq\_missile\_tests$Date[irq\_missile\_tests$Date == "May-93"] <- "1993-05-01"  
  
# Join data  
join <- dplyr::full\_join(nk\_missile\_tests, irn\_missile\_tests)

## Joining, by = c("EventID", "Date", "DateEntered", "MissileName", "MissileFamily", "FacilityName", "FacilityLatitude", "FacilityLongitude", "LandingLocation", "Apogee", "DistanceTravelled", "Confirmation", "TestOutcome", "AdditionalInformation", "Source", "Country")

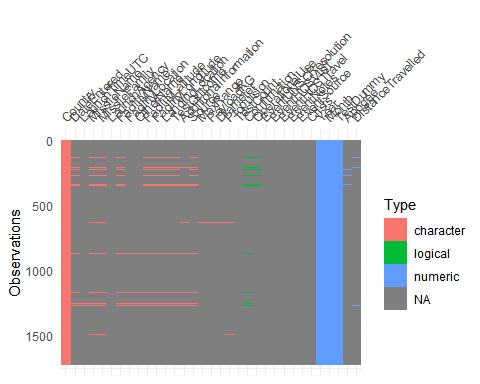
join <- dplyr::full\_join(join, irq\_missile\_tests)

## Joining, by = c("Date", "MissileName", "AdditionalInformation", "Country")

join <- dplyr::full\_join(join, pak\_missile\_tests)

## Joining, by = c("Date", "MissileName", "MissileFamily", "Country", "Propulsion")

missile\_dat\_final <- join  
rm(join)  
  
# Clean  
missile\_dat\_final$Date <- as.character(missile\_dat\_final$Date)  
missile\_dat\_final$DateEntered <- as.character(missile\_dat\_final$DateEntered)  
missile\_dat\_final$LaunchTimeUTC <- as.character(missile\_dat\_final$LaunchTimeUTC)  
  
missile\_dat\_final$MissileFamily[missile\_dat\_final$MissileName == "Scud-B"] <- "SRBM"   
missile\_dat\_final$MissileFamily[missile\_dat\_final$MissileName == "Al Hussein"] <- "SRBM"   
missile\_dat\_final$MissileFamily[missile\_dat\_final$MissileName == "Al Abbas"] <- "SRBM"   
missile\_dat\_final$MissileFamily[missile\_dat\_final$MissileName == "Condor II/ BADR-2000"] <- "MRBM"   
missile\_dat\_final$MissileFamily[missile\_dat\_final$MissileName == "FK120/ Sakr 200"] <- "SRBM"   
missile\_dat\_final$MissileFamily[missile\_dat\_final$MissileName == "Fahad (Al Fahd)"] <- "SRBM"   
missile\_dat\_final$MissileFamily[missile\_dat\_final$MissileName == "Al Abid"] <- "SLV"   
missile\_dat\_final$MissileFamily[missile\_dat\_final$MissileName == "Tammuz I"] <- "SLV"   
missile\_dat\_final$MissileFamily[missile\_dat\_final$MissileName == "Al Samoud"] <- "TBM"   
missile\_dat\_final$MissileFamily[missile\_dat\_final$MissileName == "Al Ababil"] <- "SRBM"   
missile\_dat\_final$MissileFamily[missile\_dat\_final$MissileName == "J-1"] <- "SRBM"  
  
 # Create new variables to be manually completed  
missile\_dat\_final$EventUNSCResolution <- NA  
missile\_dat\_final$EventHOSVisit <- NA  
missile\_dat\_final$EventHOSTravel <- NA  
missile\_dat\_final$EventNotes <- NA  
missile\_dat\_final$EventSource <- NA  
missile\_dat\_final$TestCount <- NA  
missile\_dat\_final$Crisis <- NA  
  
missile\_dat\_final$TestDummy <- 1  
  
 # Set Y to country/year/month  
missile\_dat\_final$Date <- stringr::str\_sub(missile\_dat\_final$Date, end = -4)  
missile\_dat\_final$Month <- stringr::str\_sub(missile\_dat\_final$Date, start = 6)  
missile\_dat\_final$Month <- as.numeric(missile\_dat\_final$Month)  
  
missile\_dat\_final$Date <- stringr::str\_sub(missile\_dat\_final$Date, end = 4)  
missile\_dat\_final <- dplyr::rename(missile\_dat\_final, Year = Date)  
missile\_dat\_final$Year = as.numeric(missile\_dat\_final$Year)   
  
missile\_dat\_final <- missile\_dat\_final %>%  
 dplyr::select(-EventID) %>%  
 dplyr::select(Country,   
 Year,   
 Month,   
 TestDummy,   
 TestCount,   
 dplyr::everything())  
  
 # Get unique missile occurances   
missile\_dat\_final <- missile\_dat\_final %>%  
 dplyr::distinct(Year, Month, Country, .keep\_all = TRUE) %>%   
 tidyr::drop\_na(Year)  
  
 # Complete years and months   
missile\_dat\_final <- dplyr::arrange(missile\_dat\_final, Country) %>%  
 dplyr::group\_by(Country)  
  
missile\_dat\_final <- missile\_dat\_final %>%  
 tidyr::complete(Country, Year = 1984:2019,   
 fill = list(incidents = 0)) %>%  
 tidyr::complete(Year, Month = 1:12,  
 fill = list(incidents = 0))  
  
 # Fill in variables  
missile\_dat\_final <- missile\_dat\_final %>%  
 dplyr::mutate(TestDummy = ifelse(is.na(TestDummy), 0, TestDummy))  
  
# Check   
visdat::vis\_dat(missile\_dat\_final)

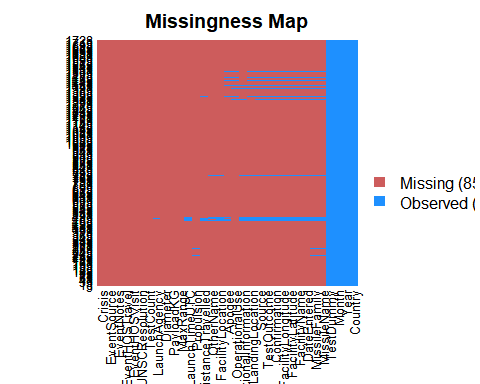


Amelia::missmap(missile\_dat\_final)

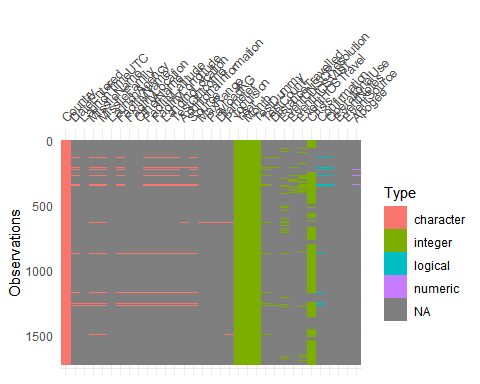
## Warning in if (class(obj) == "amelia") {: the condition has length > 1 and  
## only the first element will be used

## Warning: Unknown or uninitialised column: 'arguments'.  
  
## Warning: Unknown or uninitialised column: 'arguments'.

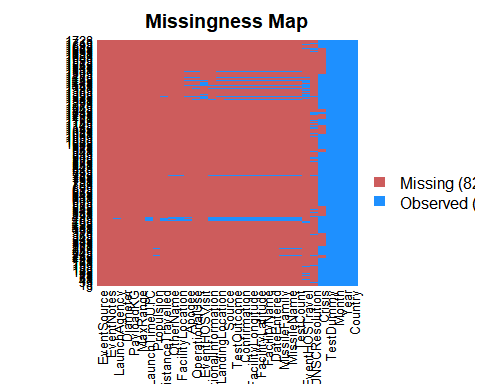
## Warning: Unknown or uninitialised column: 'imputations'.



# Save joined data  
readr::write\_csv(missile\_dat\_final, 'C:/Users/Tom Brailey/Dropbox/github\_private/MissileTest/data/missile\_dat\_final.csv')  
  
  
  
# Manaul data entry will occur at this point   
# Data will be re-uploaded into R as missile\_dat\_final\_manual\_edits  
  
  
  
# Load manually edited data  
missile\_dat\_final\_manual\_edits <- rio::import("missile\_dat\_final\_manual\_edits.csv")  
  
# Visualize data  
visdat::vis\_dat(missile\_dat\_final\_manual\_edits)



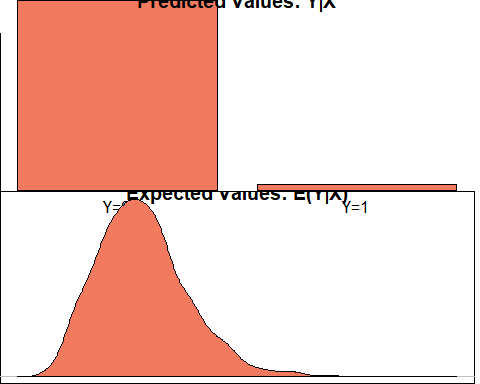
Amelia::missmap(missile\_dat\_final\_manual\_edits)



# Clean  
missile\_dat\_final\_manual\_edits$Year <- as.numeric(missile\_dat\_final\_manual\_edits$Year)  
missile\_dat\_final\_manual\_edits$Country <- as.factor(missile\_dat\_final\_manual\_edits$Country)  
missile\_dat\_final\_manual\_edits <- dplyr::as\_tibble(missile\_dat\_final\_manual\_edits)  
  
# Create lag variables for event variables  
missile\_dat\_final\_manual\_edits <- plyr::ddply(missile\_dat\_final\_manual\_edits,   
 plyr::.(Country), transform,   
 EventUNSCResolutionLag1 = c(NA, EventUNSCResolution[-length(EventUNSCResolution)])  
 )  
  
missile\_dat\_final\_manual\_edits <- plyr::ddply(missile\_dat\_final\_manual\_edits,   
 plyr::.(Country), transform,   
 EventUNSCResolutionLag2 = c(NA, EventUNSCResolutionLag1[-length(EventUNSCResolutionLag1)])  
 )  
  
# Drop columns we don't need for the analysis and organize  
missile\_dat\_final\_manual\_edits <- missile\_dat\_final\_manual\_edits %>%   
 dplyr::select(Country,  
 Year,  
 Month,  
 TestDummy,  
 TestCount,  
 MissileName,  
 MissileFamily,  
 Crisis,  
 EventUNSCResolution,  
 EventUNSCResolutionLag1,  
 EventUNSCResolutionLag2,   
 EventHOSTravel,  
 EventHOSVisit,  
 EventNotes,  
 EventSource,  
 dplyr::everything(),  
 -DateEntered)  
  
# Get rid of NA vals (for the purpose of the logit model)  
missile\_dat\_final\_manual\_edits <- missile\_dat\_final\_manual\_edits %>%  
 dplyr::mutate(EventUNSCResolution = ifelse(is.na(EventUNSCResolution), 0, EventUNSCResolution),  
 EventUNSCResolutionLag1 = ifelse(is.na(EventUNSCResolutionLag1), 0, EventUNSCResolutionLag1),  
 EventUNSCResolutionLag2 = ifelse(is.na(EventUNSCResolutionLag2), 0, EventUNSCResolutionLag2),  
 Crisis = ifelse(is.na(Crisis), 0, Crisis),  
 EventHOSTravel = ifelse(is.na(EventHOSTravel), 0, EventHOSTravel),  
 EventHOSVisit = ifelse(is.na(EventHOSVisit), 0, EventHOSVisit),  
 TestCount = ifelse(is.na(TestCount), 0, TestCount))  
  
  
  
# Set WD to save all visualizations and tables  
setwd(paste0(here::here(), '/vis/'))  
  
# Data analysis for TestDummy  
 # Logit modelling  
logit1 <- Zelig::zelig(TestDummy ~  
 Year +  
 Country +  
 EventUNSCResolution +   
 EventUNSCResolutionLag1 +   
 EventUNSCResolutionLag2 +   
 EventHOSTravel +   
 EventHOSVisit +  
 Crisis,   
 data = missile\_dat\_final\_manual\_edits,  
 model = "logit")

## How to cite this model in Zelig:  
## R Core Team. 2007.  
## logit: Logistic Regression for Dichotomous Dependent Variables  
## in Christine Choirat, Christopher Gandrud, James Honaker, Kosuke Imai, Gary King, and Olivia Lau,  
## "Zelig: Everyone's Statistical Software," http://zeligproject.org/

#stargazer::stargazer(Zelig::from\_zelig\_model(logit1),   
# type = "text",  
# title = "Logistic Regression Output (Core Hypothesis)",  
# out = "logit1.tex") # LaTeX  
#dev.off()  
  
logit1\_evs <- Zelig::setx(logit1)  
logit1\_sim <- Zelig::sim(logit1, x = logit1\_evs)  
  
logit1\_evs\_yr <- Zelig::setx(logit1, Year = c(1984:2019))  
logit1\_sim\_yr <- Zelig::sim(logit1, x = logit1\_evs\_yr)  
  
logit1\_evs\_unsc <- Zelig::setx(logit1, EventUNSCResolution = c(0:1))  
logit1\_sim\_unsc <- Zelig::sim(logit1, x = logit1\_evs\_unsc)  
  
logit1\_evs\_hosv <- Zelig::setx(logit1, EventHOSVisit = c(0:1))  
logit1\_sim\_hosv <- Zelig::sim(logit1, x = logit1\_evs\_hosv)  
  
logit1\_evs\_host <- Zelig::setx(logit1, EventHOSTravel = c(0:1))  
logit1\_sim\_host <- Zelig::sim(logit1, x = logit1\_evs\_host)  
  
logit1\_evs\_crs <- Zelig::setx(logit1, Crisis = c(0:1))  
logit1\_sim\_crs <- Zelig::sim(logit1, x = logit1\_evs\_crs)  
  
logit\_plot1 <- Zelig::plot(logit1\_sim)



jpeg(filename = "logit\_all.jpg")  
par(mfrow = c(3,2))  
logit\_plot2 <- Zelig::plot(logit1\_sim\_yr)  
logit\_plot3 <- Zelig::plot(logit1\_sim\_unsc)  
logit\_plot4 <- Zelig::plot(logit1\_sim\_hosv)  
logit\_plot5 <- Zelig::plot(logit1\_sim\_host)  
logit\_plot6 <- Zelig::plot(logit1\_sim\_crs)  
dev.off()

## png   
## 2

jpeg(filename = "logit\_by\_year.jpg")  
Zelig::plot(logit1\_sim\_yr)  
dev.off()

## png   
## 2

jpeg(filename = "logit\_summary.jpg")  
Zelig::plot(logit1\_sim)   
dev.off()

## png   
## 2

# Logit 2 modelling (ensure that I get the same results as in Zelig)  
logit2 <- glm(TestDummy ~ Year +  
 Country +  
 EventUNSCResolution +   
 EventUNSCResolutionLag1 +   
 EventUNSCResolutionLag2 +   
 EventHOSTravel +   
 EventHOSVisit +  
 Crisis,   
 data=missile\_dat\_final\_manual\_edits,   
 family=binomial(link="logit"))  
logit2\_predicted <- plogis(predict(logit2, missile\_dat\_final\_manual\_edits)) # predicted scores  
  
Zelig::summary(logit2)

##   
## Call:  
## glm(formula = TestDummy ~ Year + Country + EventUNSCResolution +   
## EventUNSCResolutionLag1 + EventUNSCResolutionLag2 + EventHOSTravel +   
## EventHOSVisit + Crisis, family = binomial(link = "logit"),   
## data = missile\_dat\_final\_manual\_edits)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.6292 -0.4031 -0.2634 -0.1263 3.1963   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.392e+02 2.314e+01 -6.015 1.80e-09 \*\*\*  
## Year 6.851e-02 1.157e-02 5.921 3.20e-09 \*\*\*  
## CountryIraq -3.509e+00 6.203e-01 -5.657 1.54e-08 \*\*\*  
## CountryNorth Korea -3.977e-01 2.525e-01 -1.575 0.1152   
## CountryPakistan -1.337e+00 3.144e-01 -4.253 2.11e-05 \*\*\*  
## EventUNSCResolution 1.411e+00 2.506e-01 5.629 1.81e-08 \*\*\*  
## EventUNSCResolutionLag1 -9.328e-02 3.112e-01 -0.300 0.7644   
## EventUNSCResolutionLag2 -2.449e-03 3.105e-01 -0.008 0.9937   
## EventHOSTravel -4.820e-01 3.052e-01 -1.579 0.1142   
## EventHOSVisit 5.837e-01 3.387e-01 1.723 0.0848 .   
## Crisis -1.202e-01 2.406e-01 -0.500 0.6172   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 922.63 on 1727 degrees of freedom  
## Residual deviance: 744.72 on 1717 degrees of freedom  
## AIC: 766.72  
##   
## Number of Fisher Scoring iterations: 7

#stargazer::stargazer(logit2,   
# type = "text",  
# title = "Logistic Regression Output (Core Hypothesis)",  
# out = "logit2.tex") # LaTeX  
#dev.off()  
  
jpeg(filename = "logit2\_coef.jpg")  
dotwhisker::dwplot(logit2)  
dev.off()

## png   
## 2

#stargazer::stargazer(logit2) # LaTeX  
  
jpeg(filename = "logit\_resid.jpg")  
par(mfrow = c(2,2))  
Zelig::plot(logit2)  
dev.off()

## png   
## 2

# Data Analysis for TestCount  
jpeg(filename = "data\_hist.jpg")  
hist(missile\_dat\_final\_manual\_edits$TestCount) # A simple histogram shows that the data contain a lot of 0s  
dev.off()

## png   
## 2

# OLS  
lm1 <- lm(TestCount ~ Year +  
 Country +  
 EventUNSCResolution +   
 EventUNSCResolutionLag1 +   
 EventUNSCResolutionLag2 +   
 EventHOSTravel +   
 EventHOSVisit +  
 Crisis,  
 data = missile\_dat\_final\_manual\_edits)  
summary(lm1)

##   
## Call:  
## lm(formula = TestCount ~ Year + Country + EventUNSCResolution +   
## EventUNSCResolutionLag1 + EventUNSCResolutionLag2 + EventHOSTravel +   
## EventHOSVisit + Crisis, data = missile\_dat\_final\_manual\_edits)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.3815 -0.2282 -0.0905 0.0183 28.9950   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -17.836814 4.657517 -3.830 0.000133 \*\*\*  
## Year 0.009053 0.002334 3.879 0.000109 \*\*\*  
## CountryIraq -0.264130 0.073355 -3.601 0.000326 \*\*\*  
## CountryNorth Korea 0.061106 0.073033 0.837 0.402882   
## CountryPakistan -0.146256 0.071344 -2.050 0.040515 \*   
## EventUNSCResolution 0.327887 0.074581 4.396 1.17e-05 \*\*\*  
## EventUNSCResolutionLag1 0.011189 0.074424 0.150 0.880514   
## EventUNSCResolutionLag2 -0.023527 0.074453 -0.316 0.752046   
## EventHOSTravel -0.211661 0.094571 -2.238 0.025340 \*   
## EventHOSVisit 0.726455 0.127469 5.699 1.41e-08 \*\*\*  
## Crisis -0.114377 0.052925 -2.161 0.030825 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.9453 on 1717 degrees of freedom  
## Multiple R-squared: 0.06179, Adjusted R-squared: 0.05633   
## F-statistic: 11.31 on 10 and 1717 DF, p-value: < 2.2e-16

lm2 <- lm(TestCount ~ Country +  
 EventUNSCResolution +   
 EventUNSCResolutionLag1 +   
 EventUNSCResolutionLag2 +   
 EventHOSTravel +   
 EventHOSVisit,  
 data = missile\_dat\_final\_manual\_edits)  
summary(lm2)

##   
## Call:  
## lm(formula = TestCount ~ Country + EventUNSCResolution + EventUNSCResolutionLag1 +   
## EventUNSCResolutionLag2 + EventHOSTravel + EventHOSVisit,   
## data = missile\_dat\_final\_manual\_edits)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.3027 -0.2414 -0.0479 0.0369 29.0462   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.192296 0.054828 3.507 0.000464 \*\*\*  
## CountryIraq -0.265710 0.072533 -3.663 0.000257 \*\*\*  
## CountryNorth Korea 0.049092 0.070237 0.699 0.484679   
## CountryPakistan -0.144389 0.069902 -2.066 0.039016 \*   
## EventUNSCResolution 0.348868 0.074735 4.668 3.28e-06 \*\*\*  
## EventUNSCResolutionLag1 0.036465 0.074489 0.490 0.624524   
## EventUNSCResolutionLag2 0.004143 0.074492 0.056 0.955657   
## EventHOSTravel -0.169106 0.092294 -1.832 0.067086 .   
## EventHOSVisit 0.761547 0.125686 6.059 1.68e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.9496 on 1719 degrees of freedom  
## Multiple R-squared: 0.05213, Adjusted R-squared: 0.04772   
## F-statistic: 11.82 on 8 and 1719 DF, p-value: < 2.2e-16

lm3 <- lm(TestCount ~ Country +  
 EventUNSCResolution +   
 EventUNSCResolutionLag1 +   
 EventUNSCResolutionLag2 +   
 EventHOSTravel +   
 EventHOSVisit +  
 Crisis:Year,  
 data = missile\_dat\_final\_manual\_edits)  
summary(lm3)

##   
## Call:  
## lm(formula = TestCount ~ Country + EventUNSCResolution + EventUNSCResolutionLag1 +   
## EventUNSCResolutionLag2 + EventHOSTravel + EventHOSVisit +   
## Crisis:Year, data = missile\_dat\_final\_manual\_edits)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.2843 -0.2285 -0.1082 0.0168 28.9800   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.297e-01 5.966e-02 3.850 0.000122 \*\*\*  
## CountryIraq -2.465e-01 7.351e-02 -3.353 0.000817 \*\*\*  
## CountryNorth Korea 8.149e-02 7.312e-02 1.114 0.265225   
## CountryPakistan -1.215e-01 7.135e-02 -1.702 0.088893 .   
## EventUNSCResolution 3.483e-01 7.470e-02 4.662 3.37e-06 \*\*\*  
## EventUNSCResolutionLag1 3.598e-02 7.446e-02 0.483 0.629037   
## EventUNSCResolutionLag2 1.734e-03 7.447e-02 0.023 0.981422   
## EventHOSTravel -1.457e-01 9.343e-02 -1.559 0.119120   
## EventHOSVisit 7.903e-01 1.269e-01 6.226 5.99e-10 \*\*\*  
## Crisis:Year -4.166e-05 2.627e-05 -1.586 0.112938   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.9492 on 1718 degrees of freedom  
## Multiple R-squared: 0.05351, Adjusted R-squared: 0.04856   
## F-statistic: 10.79 on 9 and 1718 DF, p-value: < 2.2e-16

lm4 <- lm(TestCount ~ Country +  
 EventUNSCResolution +   
 EventUNSCResolutionLag1 +   
 EventUNSCResolutionLag2 +   
 EventHOSTravel +   
 EventHOSVisit +  
 Crisis\*Year,  
 data = missile\_dat\_final\_manual\_edits)  
summary(lm4)

##   
## Call:  
## lm(formula = TestCount ~ Country + EventUNSCResolution + EventUNSCResolutionLag1 +   
## EventUNSCResolutionLag2 + EventHOSTravel + EventHOSVisit +   
## Crisis \* Year, data = missile\_dat\_final\_manual\_edits)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.3715 -0.2137 -0.0982 0.0256 28.9737   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -29.537128 11.583498 -2.550 0.010861 \*   
## CountryIraq -0.272542 0.073745 -3.696 0.000226 \*\*\*  
## CountryNorth Korea 0.036557 0.076343 0.479 0.632110   
## CountryPakistan -0.147407 0.071347 -2.066 0.038973 \*   
## EventUNSCResolution 0.328103 0.074577 4.400 1.15e-05 \*\*\*  
## EventUNSCResolutionLag1 0.012927 0.074436 0.174 0.862146   
## EventUNSCResolutionLag2 -0.021643 0.074468 -0.291 0.771368   
## EventHOSTravel -0.209207 0.094591 -2.212 0.027119 \*   
## EventHOSVisit 0.727465 0.127464 5.707 1.35e-08 \*\*\*  
## Crisis 14.085192 12.871640 1.094 0.273985   
## Year 0.014911 0.005800 2.571 0.010226 \*   
## Crisis:Year -0.007101 0.006437 -1.103 0.270105   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.9452 on 1716 degrees of freedom  
## Multiple R-squared: 0.06246, Adjusted R-squared: 0.05645   
## F-statistic: 10.39 on 11 and 1716 DF, p-value: < 2.2e-16

#stargazer::stargazer(lm1, lm2, lm3, lm4,  
# type = "text",  
# title = "Linear Model Outputs",  
# out = "lm.tex") # LaTeX  
   
 # Check for heteroscedasticity  
jpeg(filename = "het\_checks.jpg")  
par(mfrow = c(2,2))  
plot(lm1)  
dev.off()

## png   
## 2

# Robust standard errors  
lm1\_rob <- estimatr::lm\_robust(TestCount ~ Year +   
 Country +  
 EventUNSCResolution +   
 EventUNSCResolutionLag1 +   
 EventUNSCResolutionLag2 +   
 EventHOSTravel +   
 EventHOSVisit +  
 Crisis,  
 data = missile\_dat\_final\_manual\_edits)

## Warning: Assigning non-quosure objects to quosure lists is deprecated as of rlang 0.3.0.  
## Please coerce to a bare list beforehand with `as.list()`  
## This warning is displayed once per session.

summary(lm1\_rob)

##   
## Call:  
## estimatr::lm\_robust(formula = TestCount ~ Year + Country + EventUNSCResolution +   
## EventUNSCResolutionLag1 + EventUNSCResolutionLag2 + EventHOSTravel +   
## EventHOSVisit + Crisis, data = missile\_dat\_final\_manual\_edits)  
##   
## Standard error type: HC2   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|) CI Lower  
## (Intercept) -17.836814 3.531536 -5.0507 4.870e-07 -24.763381  
## Year 0.009053 0.001769 5.1171 3.450e-07 0.005583  
## CountryIraq -0.264130 0.045780 -5.7696 9.407e-09 -0.353920  
## CountryNorth Korea 0.061106 0.062054 0.9847 3.249e-01 -0.060603  
## CountryPakistan -0.146256 0.044899 -3.2575 1.146e-03 -0.234318  
## EventUNSCResolution 0.327887 0.090756 3.6128 3.116e-04 0.149882  
## EventUNSCResolutionLag1 0.011189 0.066814 0.1675 8.670e-01 -0.119857  
## EventUNSCResolutionLag2 -0.023527 0.060320 -0.3900 6.966e-01 -0.141835  
## EventHOSTravel -0.211661 0.126653 -1.6712 9.487e-02 -0.460071  
## EventHOSVisit 0.726455 0.498202 1.4582 1.450e-01 -0.250693  
## Crisis -0.114377 0.079580 -1.4373 1.508e-01 -0.270460  
## CI Upper DF  
## (Intercept) -10.91025 1717  
## Year 0.01252 1717  
## CountryIraq -0.17434 1717  
## CountryNorth Korea 0.18281 1717  
## CountryPakistan -0.05819 1717  
## EventUNSCResolution 0.50589 1717  
## EventUNSCResolutionLag1 0.14223 1717  
## EventUNSCResolutionLag2 0.09478 1717  
## EventHOSTravel 0.03675 1717  
## EventHOSVisit 1.70360 1717  
## Crisis 0.04171 1717  
##   
## Multiple R-squared: 0.06179 , Adjusted R-squared: 0.05633   
## F-statistic: 7.249 on 10 and 1717 DF, p-value: 2.636e-11

#lm1\_rob %>% # LaTeX  
# estimatr::tidy() %>%  
# xtable::xtable() %>%  
# xtable::print.xtable(file = "lm1\_rob.tex")  
  
 # Poisson modelling  
poisson1 <- Zelig::zelig(TestCount ~ Year +   
 Country +  
 EventUNSCResolution +   
 EventUNSCResolutionLag1 +   
 EventUNSCResolutionLag2 +   
 EventHOSTravel +   
 EventHOSVisit +  
 Crisis,  
 data = missile\_dat\_final\_manual\_edits,  
 model = "poisson")

## How to cite this model in Zelig:  
## R Core Team. 2007.  
## poisson: Poisson Regression for Event Count Dependent Variables  
## in Christine Choirat, Christopher Gandrud, James Honaker, Kosuke Imai, Gary King, and Olivia Lau,  
## "Zelig: Everyone's Statistical Software," http://zeligproject.org/

Zelig::summary(poisson1)

## Model:   
##   
## Call:  
## z5$zelig(formula = TestCount ~ Year + Country + EventUNSCResolution +   
## EventUNSCResolutionLag1 + EventUNSCResolutionLag2 + EventHOSTravel +   
## EventHOSVisit + Crisis, data = missile\_dat\_final\_manual\_edits)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.2061 -0.5384 -0.3250 -0.1223 12.1687   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -127.64191 15.38313 -8.298 < 2e-16  
## Year 0.06312 0.00769 8.207 2.27e-16  
## CountryIraq -3.73709 0.59317 -6.300 2.97e-10  
## CountryNorth Korea 0.04875 0.16251 0.300 0.764  
## CountryPakistan -1.24051 0.23561 -5.265 1.40e-07  
## EventUNSCResolution 1.23101 0.13742 8.958 < 2e-16  
## EventUNSCResolutionLag1 -0.01427 0.17828 -0.080 0.936  
## EventUNSCResolutionLag2 -0.25289 0.19704 -1.283 0.199  
## EventHOSTravel -0.82411 0.19689 -4.186 2.84e-05  
## EventHOSVisit 1.31293 0.18953 6.927 4.29e-12  
## Crisis -0.69555 0.13984 -4.974 6.56e-07  
##   
## (Dispersion parameter for poisson family taken to be 1)  
##   
## Null deviance: 1679.6 on 1727 degrees of freedom  
## Residual deviance: 1179.4 on 1717 degrees of freedom  
## AIC: 1516.9  
##   
## Number of Fisher Scoring iterations: 7  
##   
## Next step: Use 'setx' method

#stargazer::stargazer(Zelig::from\_zelig\_model(poisson1),  
# type = "text",  
# title = "Poisson Distribution",  
# out = "poisson1.tex") # LaTeX  
  
 # Negative binomial  
negbin1 <- Zelig::zelig(TestCount ~ Year +   
 Country +  
 EventUNSCResolution +   
 EventUNSCResolutionLag1 +   
 EventUNSCResolutionLag2 +   
 EventHOSTravel +   
 EventHOSVisit +  
 Crisis,  
 data = missile\_dat\_final\_manual\_edits,  
 model = "negbin")

## How to cite this model in Zelig:  
## William N. Venables, and Brian D. Ripley. 2008.  
## negbin: Negative Binomial Regression for Event Count Dependent Variables  
## in Christine Choirat, Christopher Gandrud, James Honaker, Kosuke Imai, Gary King, and Olivia Lau,  
## "Zelig: Everyone's Statistical Software," http://zeligproject.org/

Zelig::summary(negbin1)

## Model:   
##   
## Call:  
## z5$zelig(formula = TestCount ~ Year + Country + EventUNSCResolution +   
## EventUNSCResolutionLag1 + EventUNSCResolutionLag2 + EventHOSTravel +   
## EventHOSVisit + Crisis, data = missile\_dat\_final\_manual\_edits)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.1997 -0.4265 -0.2726 -0.1130 3.0703   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -1.605e+02 2.452e+01 -6.544 5.99e-11  
## Year 7.945e-02 1.227e-02 6.475 9.45e-11  
## CountryIraq -3.997e+00 6.559e-01 -6.095 1.10e-09  
## CountryNorth Korea -8.491e-02 2.778e-01 -0.306 0.759863  
## CountryPakistan -1.148e+00 3.204e-01 -3.583 0.000340  
## EventUNSCResolution 1.853e+00 2.914e-01 6.360 2.02e-10  
## EventUNSCResolutionLag1 7.247e-03 3.452e-01 0.021 0.983250  
## EventUNSCResolutionLag2 -4.086e-01 3.741e-01 -1.092 0.274774  
## EventHOSTravel -8.972e-01 3.642e-01 -2.463 0.013763  
## EventHOSVisit 1.325e+00 3.979e-01 3.330 0.000869  
## Crisis -7.029e-01 2.464e-01 -2.853 0.004337  
##   
## (Dispersion parameter for Negative Binomial(0.1711) family taken to be 1)  
##   
## Null deviance: 652.46 on 1727 degrees of freedom  
## Residual deviance: 404.16 on 1717 degrees of freedom  
## AIC: 1129.5  
##   
## Number of Fisher Scoring iterations: 1  
##   
##   
## Theta: 0.1711   
## Std. Err.: 0.0266   
##   
## 2 x log-likelihood: -1105.5080   
## Next step: Use 'setx' method

negbin1\_evs <- Zelig::setx(negbin1)  
negbin1\_sim <- Zelig::sim(negbin1, negbin1\_evs)  
  
negbin1\_evs\_yr <- Zelig::setx(negbin1, Year = c(1984:2019))  
negbin1\_sim\_yr <- Zelig::sim(negbin1, negbin1\_evs\_yr)  
  
negbin1\_evs\_unsc <- Zelig::setx(negbin1, EventUNSCResolution = c(0:1))  
negbin1\_sim\_unsc <- Zelig::sim(negbin1, x = negbin1\_evs\_unsc)  
  
negbin1\_evs\_hosv <- Zelig::setx(negbin1, EventHOSVisit = c(0:1))  
negbin1\_sim\_hosv <- Zelig::sim(negbin1, x = negbin1\_evs\_hosv)  
  
negbin1\_evs\_host <- Zelig::setx(negbin1, EventHOSTravel = c(0:1))  
negbin1\_sim\_host <- Zelig::sim(negbin1, x = negbin1\_evs\_host)  
  
negbin1\_evs\_crs <- Zelig::setx(negbin1, Crisis = c(0:1))  
negbin1\_sim\_crs <- Zelig::sim(negbin1, x = negbin1\_evs\_crs)  
  
jpeg(filename = "negbin\_all.jpg")  
par(mfrow = c(3,2))  
negbin\_plot2 <- Zelig::plot(negbin1\_sim\_yr)  
negbin\_plot3 <- Zelig::plot(negbin1\_sim\_unsc)  
negbin\_plot4 <- Zelig::plot(negbin1\_sim\_hosv)  
negbin\_plot5 <- Zelig::plot(negbin1\_sim\_host)  
negbin\_plot6 <- Zelig::plot(negbin1\_sim\_crs)  
dev.off()

## png   
## 2

jpeg(filename = "negbin\_summary.jpeg")  
Zelig::plot(negbin1\_sim)  
dev.off()

## png   
## 2

jpeg(filename = "negbin\_summary\_yr.jpeg")  
Zelig::plot(negbin1\_sim\_yr)  
dev.off()

## png   
## 2

jpeg(filename = "vis\_dat.jpeg")  
visdat::vis\_dat(missile\_dat\_final\_manual\_edits)  
dev.off()

## png   
## 2

#stargazer::stargazer(Zelig::from\_zelig\_model(negbin1),  
# type = "text",  
# title = "Negative Binomial Model for Count DV",  
# out = "negbin1.tex") # LaTeX  
  
  
  
# Next steps: Convert data to time-series