Reproducible research assignment week 4

Types of storms and other severe weather events with highest impact on public health and economy. US, 1950-2011

SYNOPSIS

This analysis is aimed at answering two questions: what are the types of storms and other severe weather events with highest impact on public health in the US and what are the events with highest impact on US economy, based on 1950-2011 U.S. National Oceanic and Atmospheric Administration's (NOAA) data.

The data is processed using R software and intends to follow the commonly accepted criteria for reproducible research. The plotting system ggplot2 has been used to create the figures.

Tornadoes and floods, respectively, are the answers to the above questions.

No operational recommendations are provided, but the results presented could help the subject matter experts determine the allocation of resources and the adoption of measures to mitigate the effects of severe weather events.

DATA PROCESSING

The steps followed to process the data and reach the results are presented below.

library(ggplot2)

library(dplyr)

library(readr)

library(stringr)

library(markdown)

library(knitr)

library(ggpubr)

Download the file from NOAA's website

fileurl<-"<https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2>"

download.file(fileurl, destfile = "stormdata.csv.bz2")

Import the dataset into an R object

storm<-read.csv("stormdata.csv.bz2", header = TRUE)

Check the structure of the dataset

str(storm)

In general it is a good idea as a preliminary step to convert all date variables to date format

storm<-transform(storm,BGN\_DATE=as.Date(BGN\_DATE,"%m/%d/%Y"), END\_DATE=as.Date(END\_DATE,"%m/%d/%Y"))

First objective: determine the types of storms and weather events that have highest impact on public health

It is reasonable to consider the number of fatalities as the main indicator of impact on public health. Therefore, the severity of the events will be determined based on this criterion. The number of injuries, although obviously a very important indicator, will be a secondary indicator, used to validate the conclusions reached using the first indicator, and to complement the understanding of the severity of the events.

The next step is to calculate the sum of all fatalities by type of event across the entire period under study.

stormhealth<-summarize(group\_by(storm,EVTYPE),sfat=sum(FATALITIES,na.rm=T),sinj=sum(INJURIES,na.rm=T))

The following data transformations are used to sort the types of events based on the number of fatalities, presenting the top 20 types of events individually, and grouping all other types of events that cause fewer fatalities under the general category "Other", which will simplify the presentation of the results.

stormhealth<-arrange(stormhealth,sfat,sinj)

heatoptypes<-stormhealth[(nrow(stormhealth)-20):nrow(stormhealth),]

heabottypes<-stormhealth[1:(nrow(stormhealth)-21),]

totbot<-summarize(heabottypes,sfat=sum(sfat),sinj=sum(sinj))

totbot<-mutate(totbot,EVTYPE="OTHER")

totbot<-select(totbot,EVTYPE,sfat,sinj)

heatypes<-rbind(totbot ,heatoptypes)

The last table obtained above, "heatypes", contains the data re-arranged in a list with the 20 events that have caused more fatalities. It also presents a 21st group with the sum of the fatalities caused by all other events. These results are shown below, in the corresponding figure in the results section.

Second objective: determine the types of weather events with highest economic impact.

The NOAA data available show two different measures, the impact on property and the impact on crops. It seems reasonable in this case to add both types of economic losses in order to determine the event with highest impact.

As explained in NOAA's documentation, the table reports the economic losses in a variable that contains a number with three significant digits, and the units in which that number is expressed (K usd, M usd, B usd) are stored in another variable. This applies both to property and crops losses. Therefore the first step in the analysis is to standardize the economic losses as a number expressed in US Dollars.

The following lines of code convert the units variables to text format, then multiplies the three digit number by the corresponding factor (10^3, 10^6, 10^9). They do so for property and crops losses.

stormunits<-transform(storm, PROPDMGEXP=as.character(PROPDMGEXP), CROPDMGEXP=as.character(CROPDMGEXP),PROPDMG=ifelse(storm$PROPDMGEXP=="K",PROPDMG\*1000, ifelse(storm$PROPDMGEXP=="M",PROPDMG\*10^6, ifelse(storm$PROPDMGEXP=="B",PROPDMG\*10^9,PROPDMG))),

CROPDMG=ifelse(storm$CROPDMGEXP=="K",CROPDMG\*1000, ifelse(storm$CROPDMGEXP=="M",CROPDMG\*10^6, ifelse(storm$CROPDMGEXP=="B",CROPDMG\*10^9,CROPDMG))))

The next step is to calculate the sum of all losses by type of event across the entire period under study. In the same line of code a new variable is created, "totdmg", that is the sum of the property and crops losses in USD by type of event during the period in study.

stormdmg<-summarize(group\_by(stormunits,EVTYPE),sprop =sum(PROPDMG,na.rm=T),scrop=sum(CROPDMG,na.rm=T),totdmg=sprop+scrop)

Just like in the case of the first question, the following data transformations are used to sort the types of events based on the economic losses, presenting the top 20 types of events individually, and grouping all other types of events that cause fewer fatalities under the general category "Other", which will simplify the presentation of the results.

stormdmg<-arrange(stormdmg,totdmg,sprop,scrop)

ecotoptypes<-stormdmg[(nrow(stormdmg)-20):nrow(stormdmg),]

ecobottypes<-stormdmg[1:(nrow(stormdmg)-21),]

totbot<-summarize(ecobottypes,sprop=sum(sprop),scrop=sum(scrop),totdmg=sprop+scrop)

totbot<-mutate(totbot,EVTYPE="OTHER")

totbot<-select(totbot,EVTYPE,sprop,scrop,totdmg)

ecotypes<-rbind(totbot,ecotoptypes)

The last table above, "ecotypes", contains the data re-arranged in a list of the 20 events that have caused more fatalities. It also presents a 21st group that shows the sum of the fatalities caused by all other events. The results are presented in the corresponding figure in the results section below.

RESULTS

Impact on public health

The code below is used to construct a figure that contains two charts: the first chart on the left shows that Tornadoes are the type of weather event that have caused more fatalities in the US in the period under review, followed by Excessive heat and Flash flood. Following the criterion explained above, Tornadoes are the type of weather event with highest impact on public health in the period under review.

The second chart in the figure shows that Tornadoes are also the most impactful type of event in terms of personal injuries caused.

##dev.new()

wrapper <- function(x, ...)

{

paste(strwrap(x, ...), collapse = "\n")

}

main\_title <-**"Number of fatalities caused by storms and other weather events between 1950 and 2011 in the US. Source: NOAA"**

g<-ggplot(heatypes,aes(x=factor(EVTYPE,level=EVTYPE),y=sfat))+geom\_col(aes(fill=EVTYPE))+theme\_bw()+theme(**axis.text.x = element\_text(size=8,angle=85,vjust=0.6),axis.text.y=element\_text(size=8)**)+theme(axis.line = element\_line(colour = "blue"), panel.border = element\_blank())+theme(legend.position="none"**)+labs(x="Event Type",y="Fatalities",size=9)**+ ggtitle(wrapper(main\_title, width = 50))+scale\_y\_continuous(labels=scales::comma\_format(accuracy=1), expand = c(0, 0),breaks= c(seq(0,max(heatypes$sfat),by=1000) , max(heatypes$sfat)))+geom\_hline(yintercept=max(heatypes$sfat),linetype="dashed",color="violet")+coord\_flip()

##print(g)

##dev.new()

inj\_title <-**"Number of injuries caused by storms and weather other events between 1950 and 2011 in the US – Source: NOAA"**

h<-ggplot(heatypes,aes(x=factor(EVTYPE,level=EVTYPE),y=sinj))+geom\_col(aes(fill=EVTYPE))+theme\_bw()+theme(**axis.text.x = element\_text(size=8,angle=85,vjust=0.6),axis.text.y=element\_text(size=8)**)+theme(axis.line = element\_line(colour = "blue"), panel.border = element\_blank())+theme(legend.position="none"**)+labs(x=" ",y="Injuries",size=9)**+ ggtitle(wrapper(inj\_title, width = 50))+scale\_y\_continuous(labels=scales::comma\_format(accuracy=1), expand = c(0, 0),breaks= c(seq(0,max(heatypes$sinj),by=20000) , max(heatypes$sinj)))+geom\_hline(yintercept=max(heatypes$sinj),linetype="dashed",color="violet")+coord\_flip()

##print(h)

health<-ggarrange(g,h,labels=c("", ""),ncol = 2, nrow = 1)

print(health)

Impact on the economy

The code below is used to construct a figure that contains three charts: the first chart on the left shows that Floods are the type of weather event that have caused highest economic losses in the period under review combining property and crops losses. Floods are followed by Hurricanes/Typhoons and Tornadoes. Therefore, and consistent with the criterion chosen, Floods are the type of weather event with highest economic impact in the US in the period under review.

The second chart in the figure shows that Floods are also the most impactful type of event in terms of property damage. An interesting fact shown by the third chart is that the economic impact on crops is significantly lower than the impact on property. The three charts on the figure maintain the same horizontal scale (Billion USD) to make this fact evident.

wrapper <- function(x, ...)

{

paste(strwrap(x, ...), collapse = "\n")

}

##pdf("total\_impact.pdf", width=4, height=6)

##dev.new()

main\_title <-**"Total economic impact of storms and other weather events between 1950 and 2011 in the US. Source: NOAA"**

p<-ggplot(ecotypes,aes(x=factor(EVTYPE,level=EVTYPE),y=totdmg))+geom\_col(aes(fill=EVTYPE))+theme\_bw()+theme(**axis.text.x = element\_text(size=9,angle=85,vjust=0.6),axis.text.y=element\_text(size=8)**)+ theme(axis.line = element\_line(colour = "blue"), panel.border = element\_blank())+theme(legend.position="none"**)+labs(x="Event Type",y="BILLION USD",size=6)**+ theme(plot.title = element\_text(size=12))+ggtitle(wrapper(main\_title, width=25))+scale\_y\_continuous(limits=c(0,151\*10^9),labels=function(x)x/10^9,expand =c(0,0))+geom\_hline(yintercept=max(ecotypes$totdmg),linetype="dashed",color="violet")+coord\_flip()

##print(p)

##dev.copy2pdf(file="total\_impact.pdf",width=20, height=40)

##dev.off()

##dev.new()

prop\_title<-"**Economic impact on property of storms and other weather events between 1950 and 2011 in the US. Source: NOAA "**

m<-ggplot(ecotypes,aes(x=factor(EVTYPE,level=EVTYPE),y=sprop))+geom\_col(aes(fill=EVTYPE))+ theme\_bw()+theme(**axis.text.x = element\_text(size=9,angle=85,vjust=0.6),axis.text.y=element\_text(size=8)**)+theme(axis.line = element\_line(colour = "blue"), panel.border = element\_blank())+theme(legend.position="none"**)+labs(x=" ",y="BILLION USD",size=6)**+ theme(plot.title = element\_text(size=12))+ggtitle(wrapper(prop\_title, width = 25))+scale\_y\_continuous(limits=c(0,150\*10^9),labels=function(x)x/10^9, expand = c(0, 0))+geom\_hline(yintercept=max(ecotypes$sprop),linetype="dashed",color="violet")+coord\_flip()

##print(m)

##dev.new()

crop\_title<-"**Economic impact on crops of storms and other weather events between 1950 and 2011 in the US. Source: NOAA "**

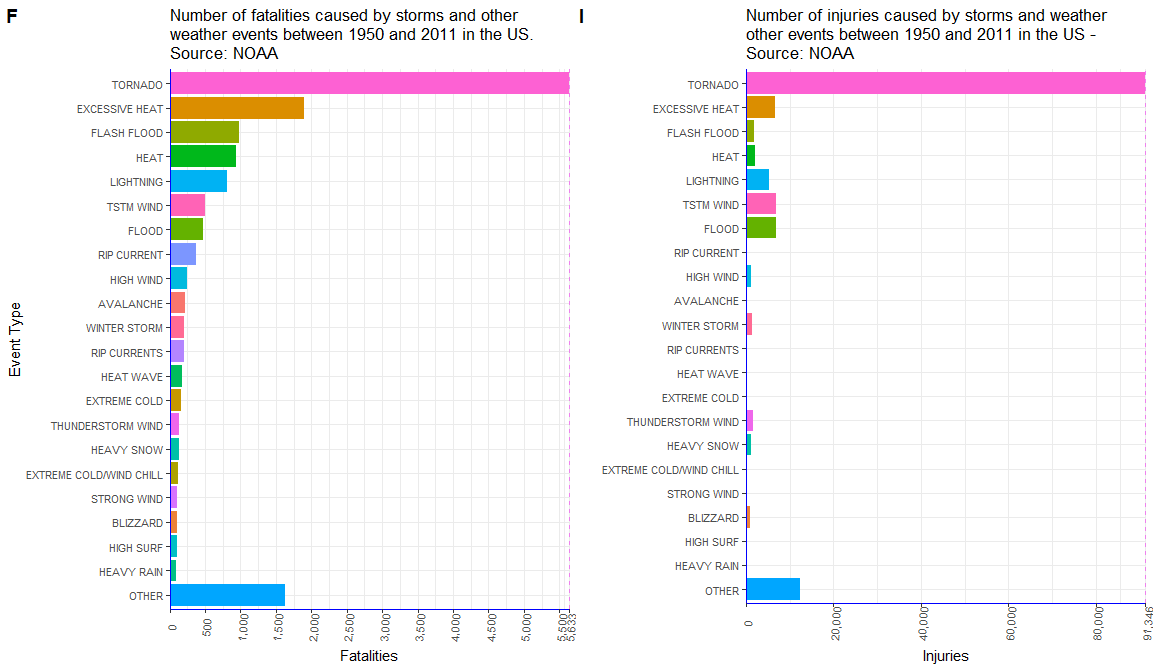
n<-ggplot(ecotypes,aes(x=factor(EVTYPE,level=EVTYPE),y=scrop))+geom\_col(aes(fill=EVTYPE))+ theme\_bw()+theme(**axis.text.x = element\_text(size=9,angle=85,vjust=0.6),axis.text.y=element\_text(size=8)**)+theme(axis.line = element\_line(colour = "blue"), panel.border = element\_blank())+theme(legend.position="none"**)+labs(x=" ",y="BILLION USD",size=6)**+ theme(plot.title = element\_text(size=12))+ggtitle(wrapper(crop\_title, width = 25))+scale\_y\_continuous(limits=c(0,150\*10^9),labels=function(x)x/10^9, expand = c(0, 0))+geom\_hline(yintercept=max(ecotypes$scrop),linetype="dashed",color="violet")+coord\_flip()

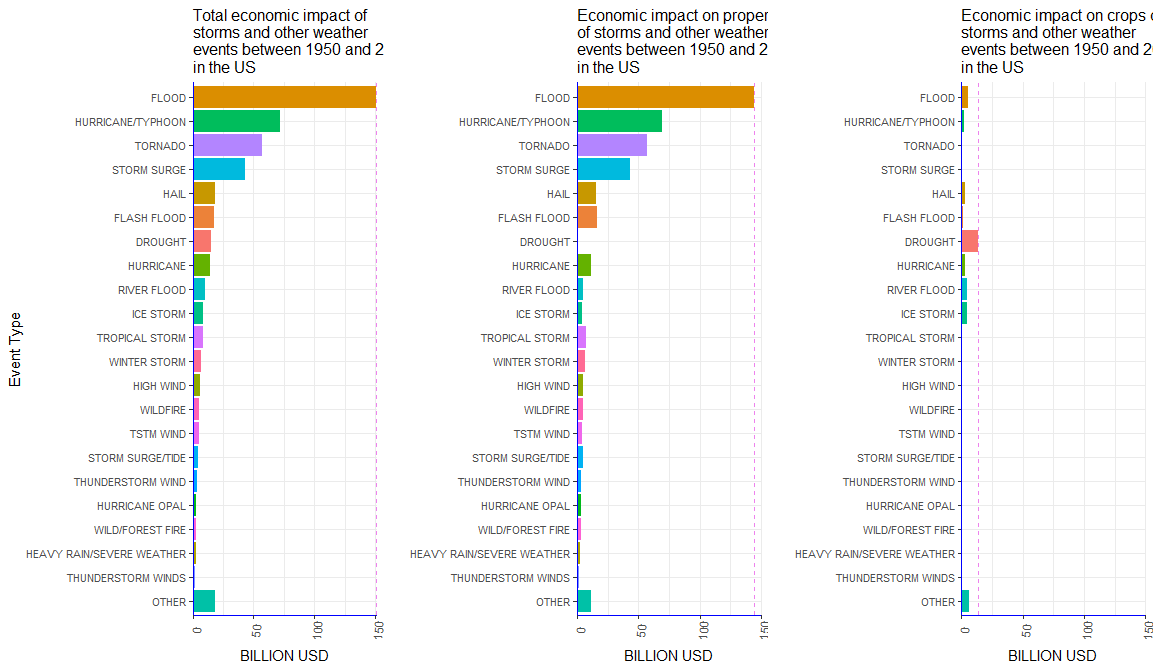
##print(n)

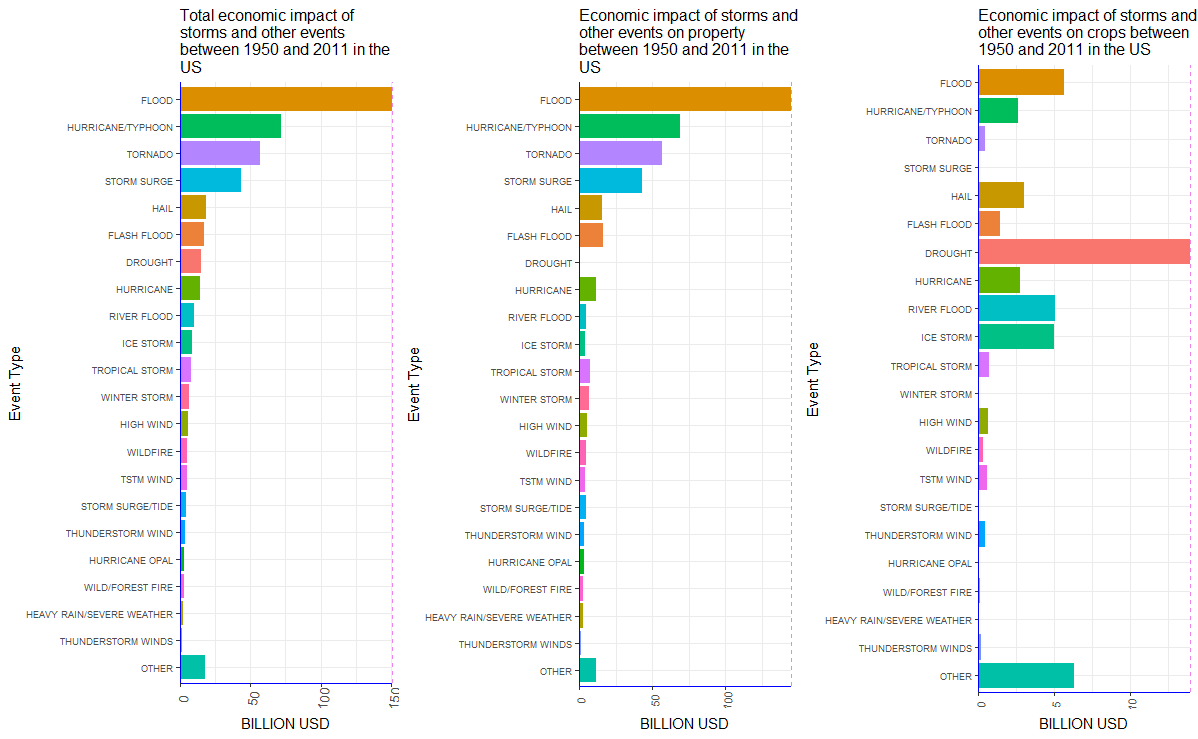
economic<-ggarrange(p,m,n,labels=c("", "",""),ncol = 3, nrow = 1)

print(economic)

<http://rpubs.com/thecmos/repres_storms_1>







##storm<-read.csv("stormdata.csv.bz2", header = TRUE, sep = ",", quote = "\"",...)

stormdmg2<-transform(stormdmg,scrop= formatC**(**as.numeric**(scrop)**,format="f",digits=0,big.mark=","**),sprop=** formatC**(**as.numeric**(sprop)**,format="f",digits=0,big.mark=","**),totdmg=** formatC**(**as.numeric**(totdmg)**,format="f",digits=0,big.mark=","**))**

ecotypes2<- transform(ecotypes,scrop= formatC**(**as.numeric**(scrop)**,format="f",digits=0,big.mark=","**),sprop=** formatC**(**as.numeric**(sprop)**,format="f",digits=0,big.mark=","**),totdmg=** formatC**(**as.numeric**(totdmg)**,format="f",digits=0,big.mark=","**))**

wrapper <- function(x, ...)

{

paste(strwrap(x, ...), collapse = "\n")

}

##pdf("total\_impact.pdf", width=4, height=6)

##dev.new()

main\_title <-**"Total economic impact of storms and other weather events between 1950 and 2011 in the US. Source: NOAA"**

p<-ggplot(ecotypes,aes(x=factor(EVTYPE,level=EVTYPE),y=totdmg))+geom\_col(aes(fill=EVTYPE))+theme\_bw()+theme(**axis.text.x = element\_text(size=9,angle=85,vjust=0.6),axis.text.y=element\_text(size=8)**)+ theme(axis.line = element\_line(colour = "blue"), panel.border = element\_blank())+theme(legend.position="none"**)+labs(x="Event Type",y="Total impact - Billion USD",size=6)**+ theme(plot.title = element\_text(size=12))+ scale\_y\_continuous(limits=c(0,151\*10^9),labels=function(x)x/10^9,expand =c(0,0))+geom\_hline(yintercept=max(ecotypes$totdmg),linetype="dashed",color="violet")+coord\_flip()

##print(p)

##dev.copy2pdf(file="total\_impact.pdf",width=20, height=40)

##dev.off()

##dev.new()

prop\_title<-"**Economic impact on property of storms and other weather events between 1950 and 2011 in the US. Source: NOAA "**

m<-ggplot(ecotypes,aes(x=factor(EVTYPE,level=EVTYPE),y=sprop))+geom\_col(aes(fill=EVTYPE))+ theme\_bw()+theme(axis.title.y=element\_blank(),axis.text.y=element\_blank(),axis.ticks.y=element\_blank()**,axis.text.x =element\_text(size=9,angle=85,vjust=0.6)**)+theme(axis.line = element\_line(colour = "blue"), panel.border = element\_blank())+theme(legend.position="none"**)+labs(x=" ",y="Impact on property - Billion USD",size=6)**+ theme(plot.title = element\_text(size=12))+scale\_y\_continuous(limits=c(0,150\*10^9),labels=function(x)x/10^9, expand = c(0, 0))+geom\_hline(yintercept=max(ecotypes$sprop),linetype="dashed",color="violet")+coord\_flip()

##print(m)

##dev.new()

crop\_title<-"**Economic impact on crops of storms and other weather events between 1950 and 2011 in the US. Source: NOAA "**

n<-ggplot(ecotypes,aes(x=factor(EVTYPE,level=EVTYPE),y=scrop))+geom\_col(aes(fill=EVTYPE))+ theme\_bw()+theme(axis.title.y=element\_blank(),axis.text.y=element\_blank(),axis.ticks.y=element\_blank()**,axis.text.x =element\_text(size=9,angle=85,vjust=0.6)**)+theme(axis.line = element\_line(colour = "blue"), panel.border = element\_blank())+theme(legend.position="none"**)+labs(x=" ",y="Impact on crops - Billion USD",size=6)**+scale\_y\_continuous(limits=c(0,150\*10^9),labels=function(x)x/10^9, expand = c(0, 0))+geom\_hline(yintercept=max(ecotypes$scrop),linetype="dashed",color="violet")+coord\_flip()

##print(n)

economic<-ggarrange(p,m,n,labels=c("Economic impact of weather events in the US. Source: NOAA", "",""),ncol = 3, nrow = 1)

print(economic)