Reproducible research assignment week 4

Types of storms and other severe weather events with highest impact on public health and economy in the US, based on 1950-2011 U.S. National Oceanic and Atmospheric Administration's (NOAA) data

##Download the file from NOAA's website

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

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

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

##Read the file into R

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

##Check the structure of the file

str(storm)

##I believe it will be a good idea to convert all dates to date format, and also, even if I am not sure this will be useful, I will create another variable, EVTYPECH, that will be the conversion of the original EVTYPE to character.

storm$EVTYPECH<-as.character(storm$EVTYPE)

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

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

> max(storm$END\_DATE,na.rm=T)

[1] "2011-11-30"

> max(storm$BGN\_DATE,na.rm=T)

[1] "2011-11-30"

stormhealth<-arrange(stormhealth,desc(sfat),desc(sinj))

##head(stormhealth)

# A tibble: 6 x 3

# EVTYPE sfat sinj

# <fct> <dbl> <dbl>

#1 TORNADO 5633 91346

#2 EXCESSIVE HEAT 1903 6525

#3 FLASH FLOOD 978 1777

#4 HEAT 937 2100

#5 LIGHTNING 816 5230

#6 TSTM WIND 504 6957

heatoptypes<-stormhealth[1:20,]

heabottypes<-stormhealth[21:985,]

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

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

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

heatypes<-rbind(heatoptypes,totbot)

wrapper <- function(x, ...)

{

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

}

main\_title <-**"Health impact of storms and other events between 1950 and 2011 in the US"**

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$sfat),by=500) , max(heatypes$sfat)))+geom\_hline(yintercept=max(heatypes$sfat),linetype="dashed",color="violet")+coord\_flip()

print(g)

inj\_title <-**"Health impact of storms and other events between 1950 and 2011 in the US"**

g<-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="Event Type",y="Injured people",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(g)

#> filter(stormhealth,sfat==max(sfat))

# A tibble: 1 x 3

# EVTYPE sfat sinj

# <fct> <dbl> <dbl>

#1 TORNADO 5633 91346

#> filter(stormhealth,sinj==max(sinj))

# A tibble: 1 x 3

# EVTYPE sfat sinj

# <fct> <dbl> <dbl>

#1 TORNADO 5633 91346

WATCHOUT,starting on row

368798

The dataset is a mess for abot 2k rows

**##ECONOMIC IMPACT**

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))))

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

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

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=","**))**

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)

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=","**))**

##head(ecotypes2)

## EVTYPE sprop scrop totdmg

##1 FLOOD 144,657,709,807 5,661,968,450 150,319,678,257

##2 HURRICANE/TYPHOON 69,305,840,000 2,607,872,800 71,913,712,800

##3 TORNADO 56,925,660,790 414,953,270 57,340,614,060

##4 STORM SURGE 43,323,536,000 5,000 43,323,541,000

##5 HAIL 15,727,367,053 3,025,537,890 18,752,904,943

##6 FLASH FLOOD 16,140,812,067 1,421,317,100 17,562,129,167

wrapper <- function(x, ...)

{

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

}

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

main\_title <-**"Economic impact of storms and other events between 1950 and 2011 in the US"**

h<-ggplot(ecotypes,aes(x=factor(EVTYPE,level=EVTYPE),y=totdmg))+geom\_col(aes(fill=EVTYPE))+theme\_bw()+theme(**axis.text.x = element\_text(size=10,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="Million USD",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(ecotypes$totdmg),by=10\*10^9) , max(ecotypes$totdmg)))+geom\_hline(yintercept=max(ecotypes$totdmg),linetype="dashed",color="violet")+coord\_flip()

print(h)

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

##dev.off()

dev.new()

prop\_title<-" **Economic impact of storms and other events on property between 1950 and 2011 in the US (usd)"**

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=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="USD",size=9)**+ ggtitle(wrapper(prop\_title, width = 50))+scale\_y\_continuous(labels=scales::comma\_format(accuracy=1), expand = c(0, 0),breaks= c(seq(0,max(ecotypes$sprop),by=500000) , max(ecotypes$sprop)))+geom\_hline(yintercept=max(ecotypes$sprop),linetype="dashed",color="violet")+coord\_flip()

print(m)

dev.new()

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

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=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="USD",size=9)**+ ggtitle(wrapper(crop\_title, width = 50))+scale\_y\_continuous(labels=scales::comma\_format(accuracy=1), expand = c(0, 0),breaks= c(seq(0,max(ecotypes$scrop),by=100000) , max(ecotypes$scrop)))+geom\_hline(yintercept=max(ecotypes$scrop),linetype="dashed",color="violet")+coord\_flip()

print(n)