Mapping Hurricane Sandy Housing Damage in New York City

Charles DiMaggio

Center for Injury Epidemiology and Prevention at Columbia
Columbia University Departments of Anesthesiology and Epidemiology
cjd11@columbia.edu
www.columbia.edu/~cjd11/charles_dimaggio/DIRE/

July 10, 2013

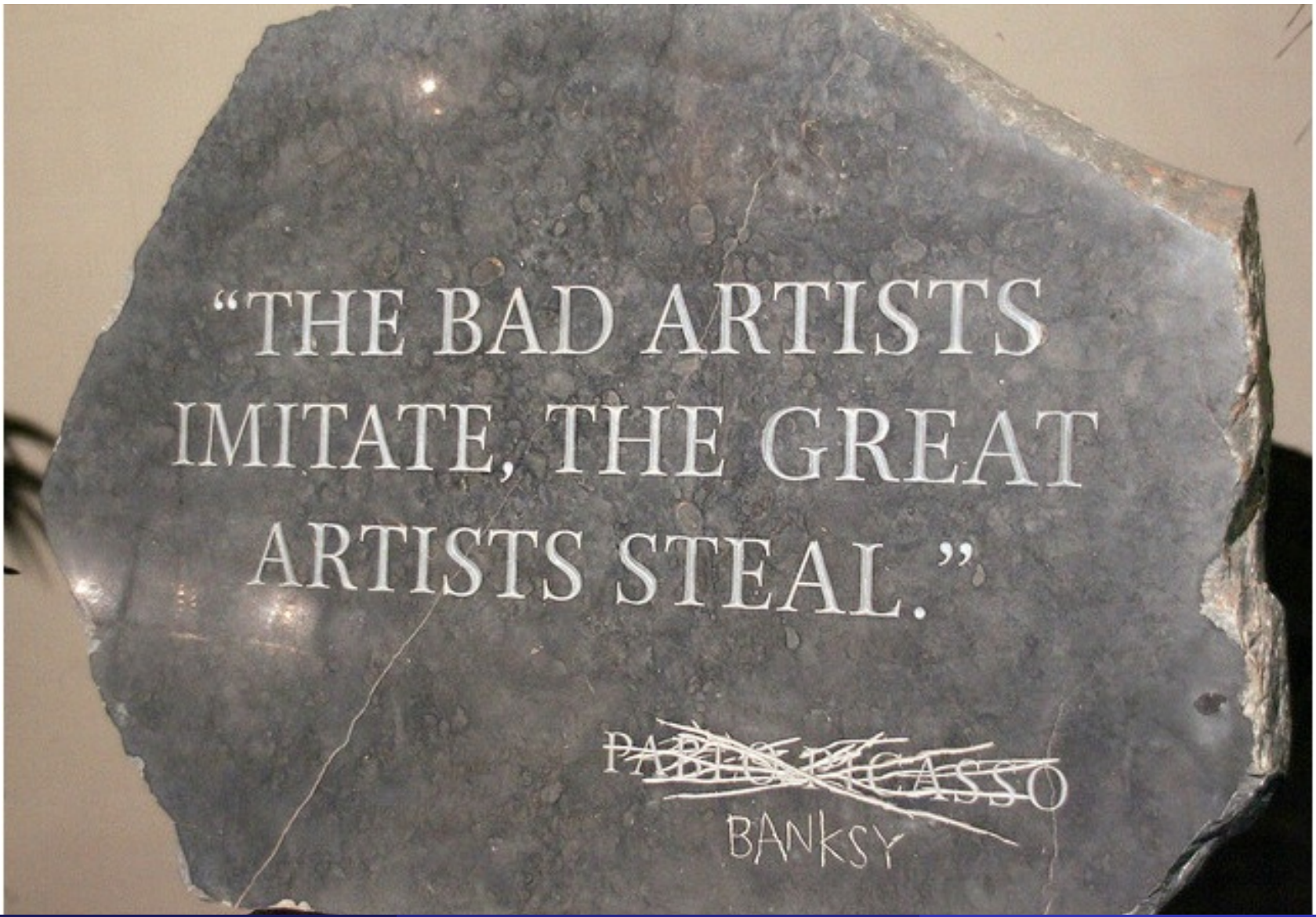
Outline

conclusions

- it's all about the data
  - just because you *can* map, doesn't mean you *should* map
- post-disaster surveillance a good place for maps
- map in R
- buy Bivand, Prebesma and Gomez-Rubio
- use ggplot

Acknowledgements

Steal from the best



stolen shamelessly from...

- Bivand, Prebesma and Gomez-Rubio
- David Kahle
- others too numerous to mention

why map?

- text for a few numbers
- table for many numbers
- plots for relationships
- maps for patterns in space and time. . .

The greatest map?

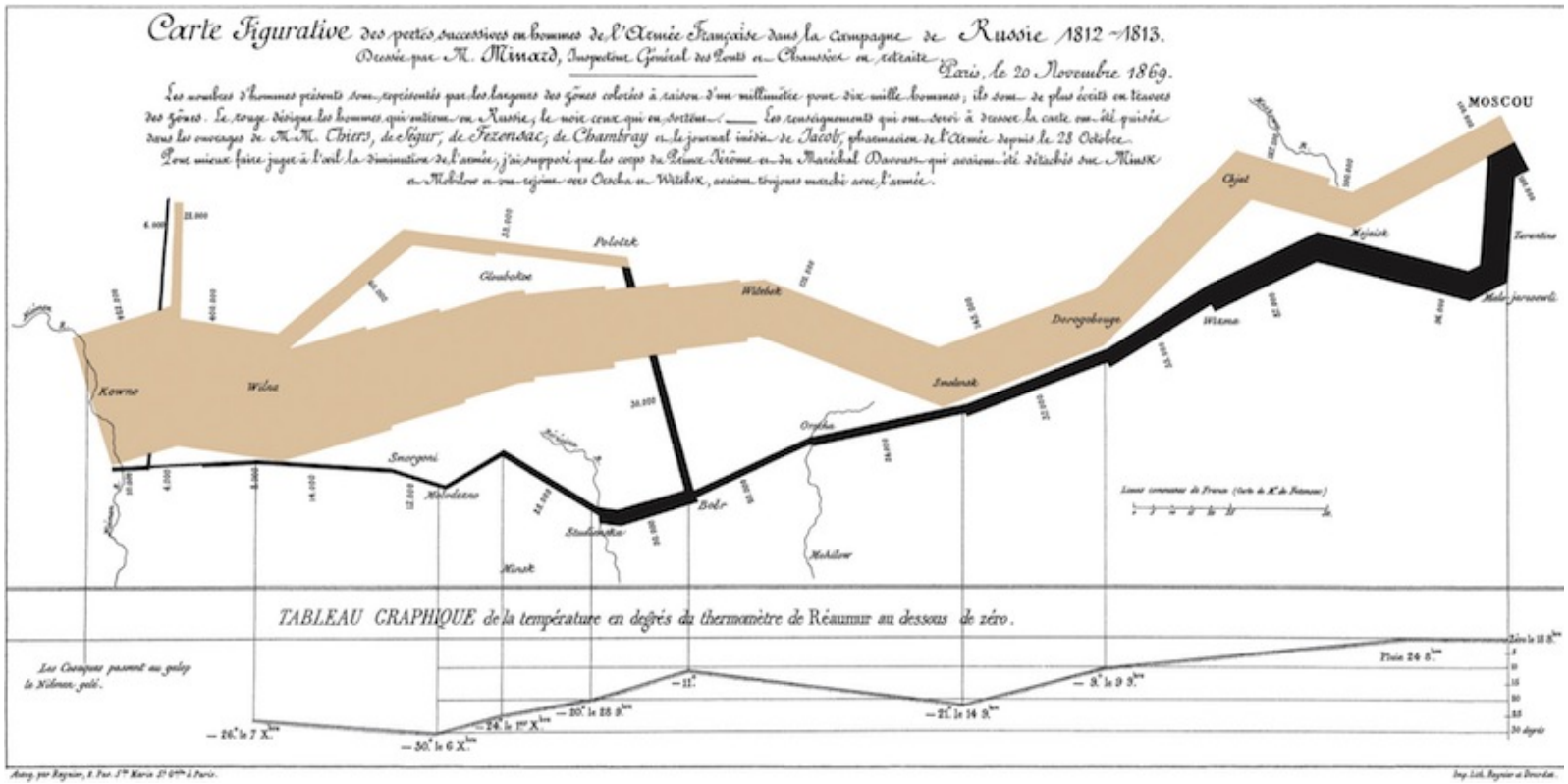


Figure : Minard's Map of Napolean's Russian Campaign of 1812

Edward Tufte would say yes...

1. comparisons
   - pre-post troop levels illustrated at beginning and end (started with 400K, ended with 10K)
2. connections
   - path of retreating army tied to a temperature scale at bottom
3. complexity
   - 6 dimensions: army size, location, direction, temperature, places
4. cleverness (use ‘whatever it takes’ to explain)
   - map annotated all over with numbers and words
5. credibility
   - two paragraphs document sourcing and detail
6. content is king - essentially an anti-war poster

why map disasters?

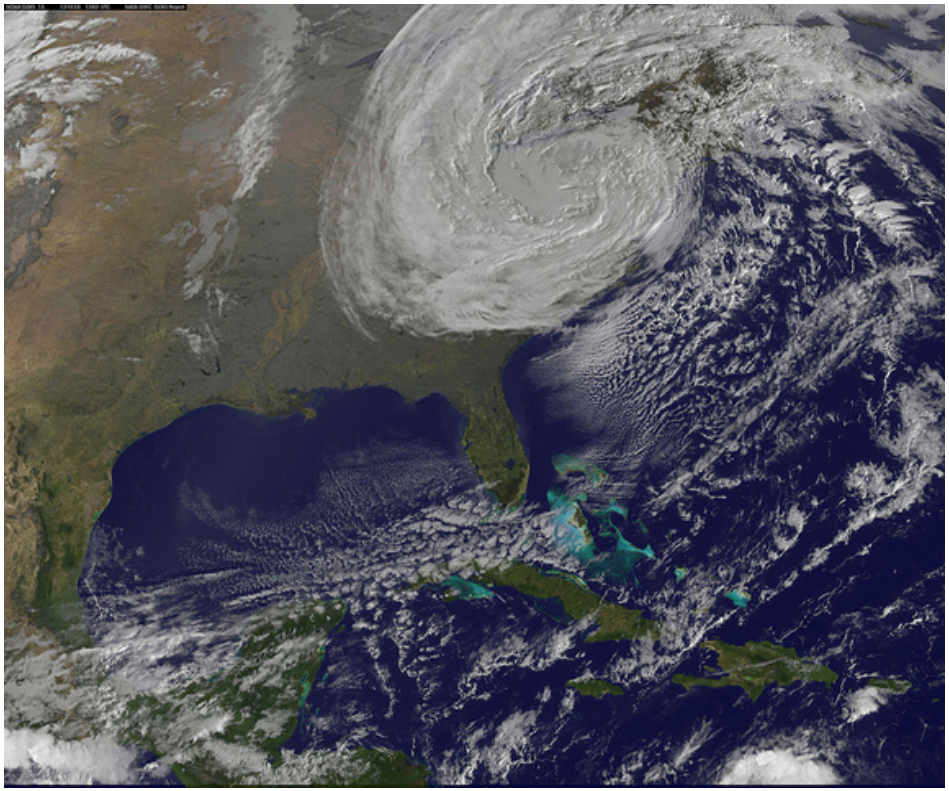
- data to make informed decisions
  - epidemiological surveillance - the ongoing systematic collection, analysis, interpretation and dissemination of health data
  - '...the sacrifice in promptness required to collect the information necessary to provide apt and well-directed aid is more than justified by the improved results'
- traditionally observation and surveys
- in place prior to disaster, active vs. passive
- but, damaged infrastructure, no uniform definitions, multiple conflicting sources
- non-traditional - drop-in, syndromic (ED, pharmacy fills), relief workers, newspaper accounts, *spatial*

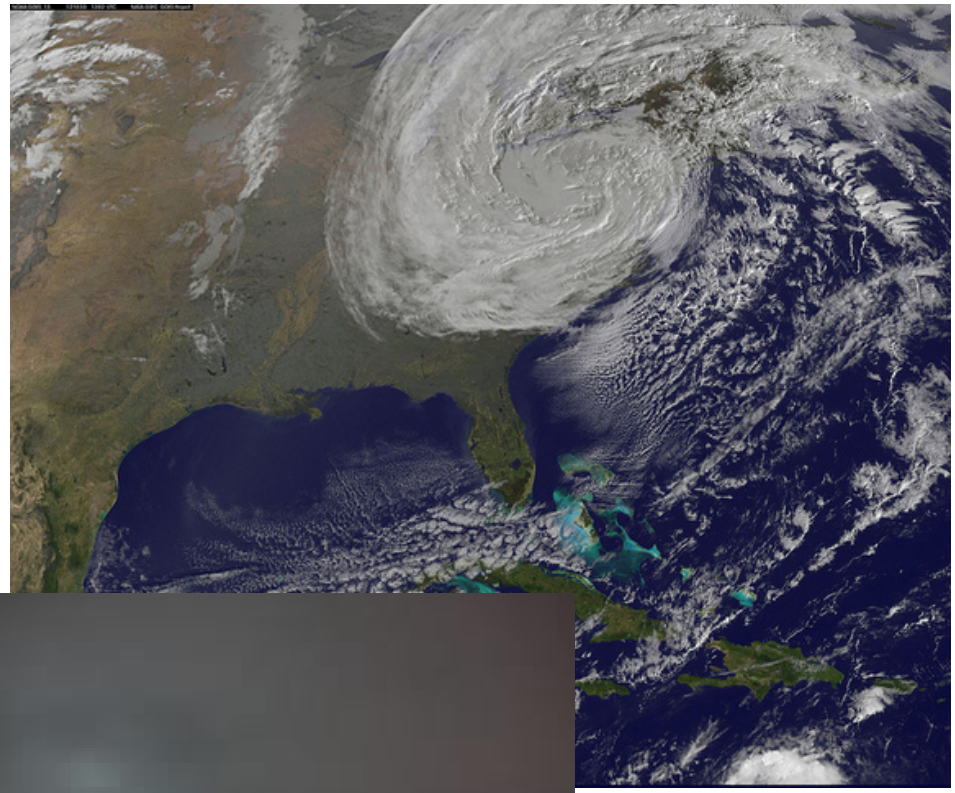
why map in R?

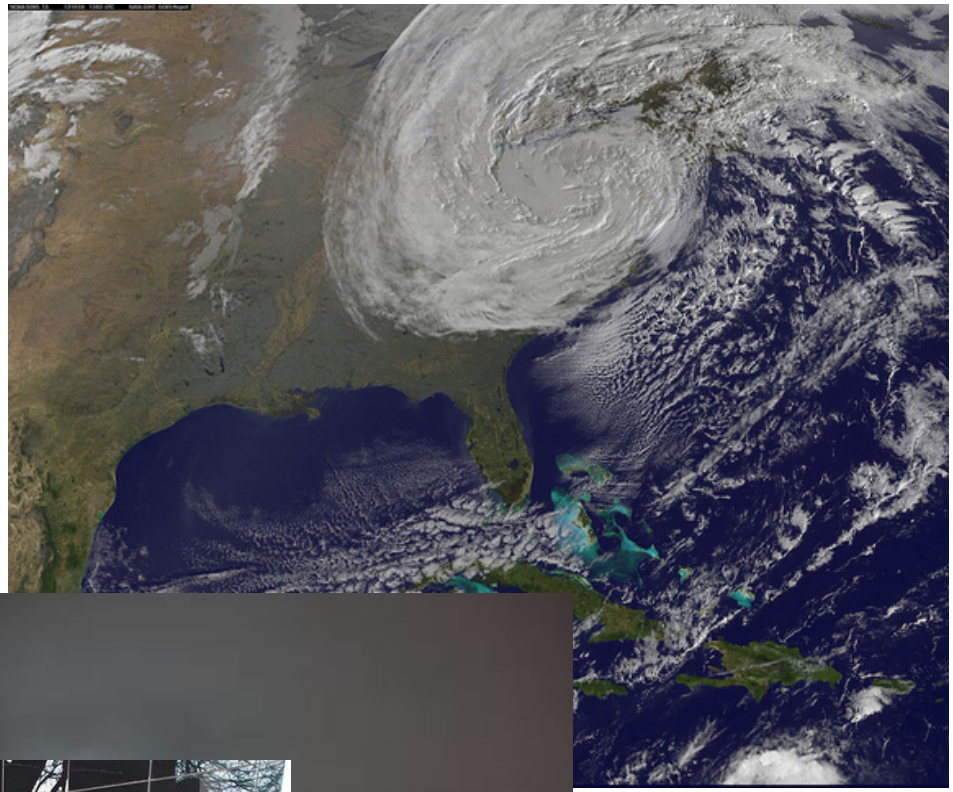
- familiar, unifying analytic environment
- many specialized packages and functions available
  - ability to translate most any spatial data
- rapidly incorporate advances
- growing, almost invariably supportive community

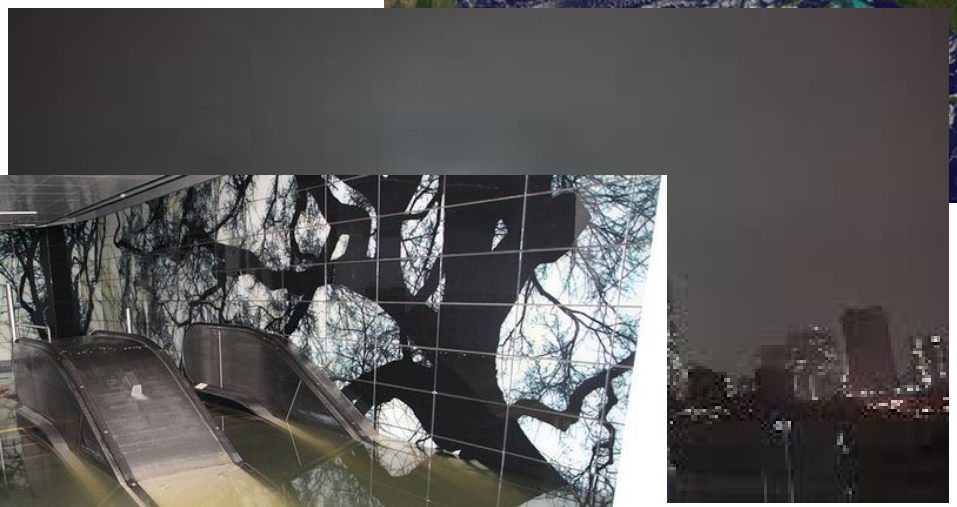
Sandy

- Oct 29-30, 2012
- most devastating cat 1 hurricane to hit eastern US in recorded history (pressure -940 mb)
- flood surges 13 ft. (4 1/3km), sustained winds 80 mph (133 kph)
- tunnels and subways flooded, city paralyzed, hospitals evacuated, 12,000 flights cancelled, JFK, LGA, Newark grounded for nearly a week
- 4.8 million without power (some for weeks)
- est $50 Billion (*E*39 Billion) damage
- 100 homes Breezy Pt. Queens burned to the ground
- 3.5 million Sandy hashtag tweets

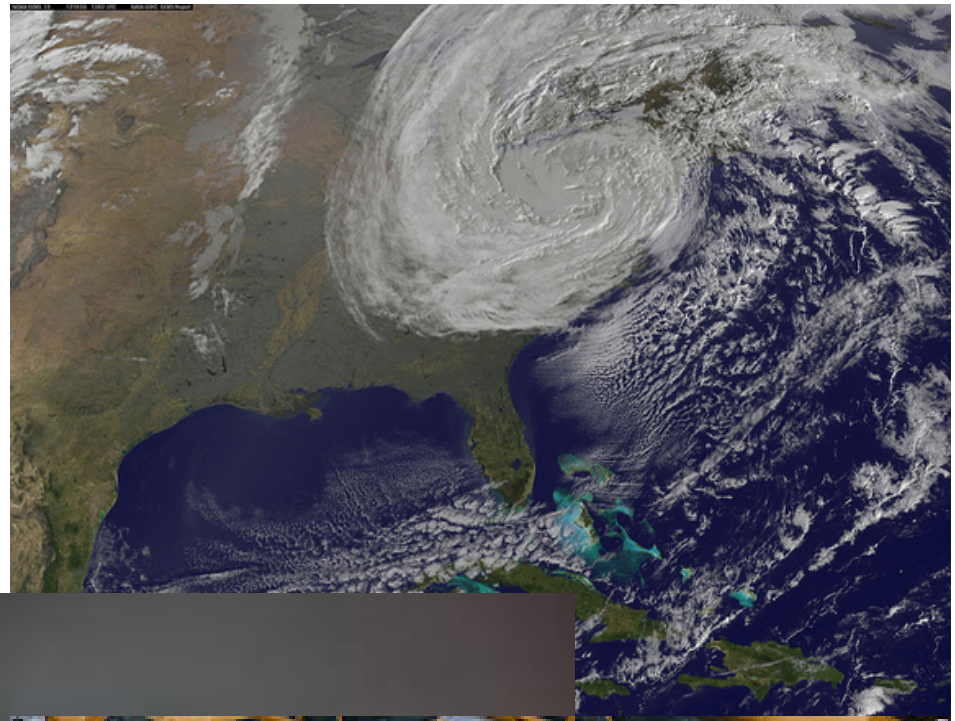


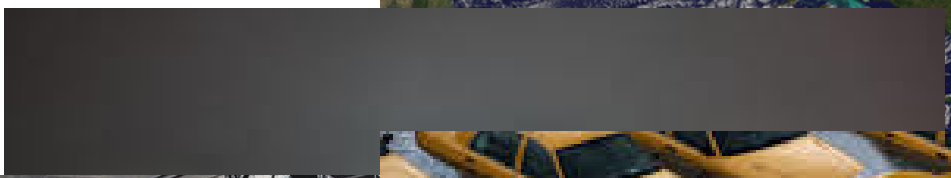




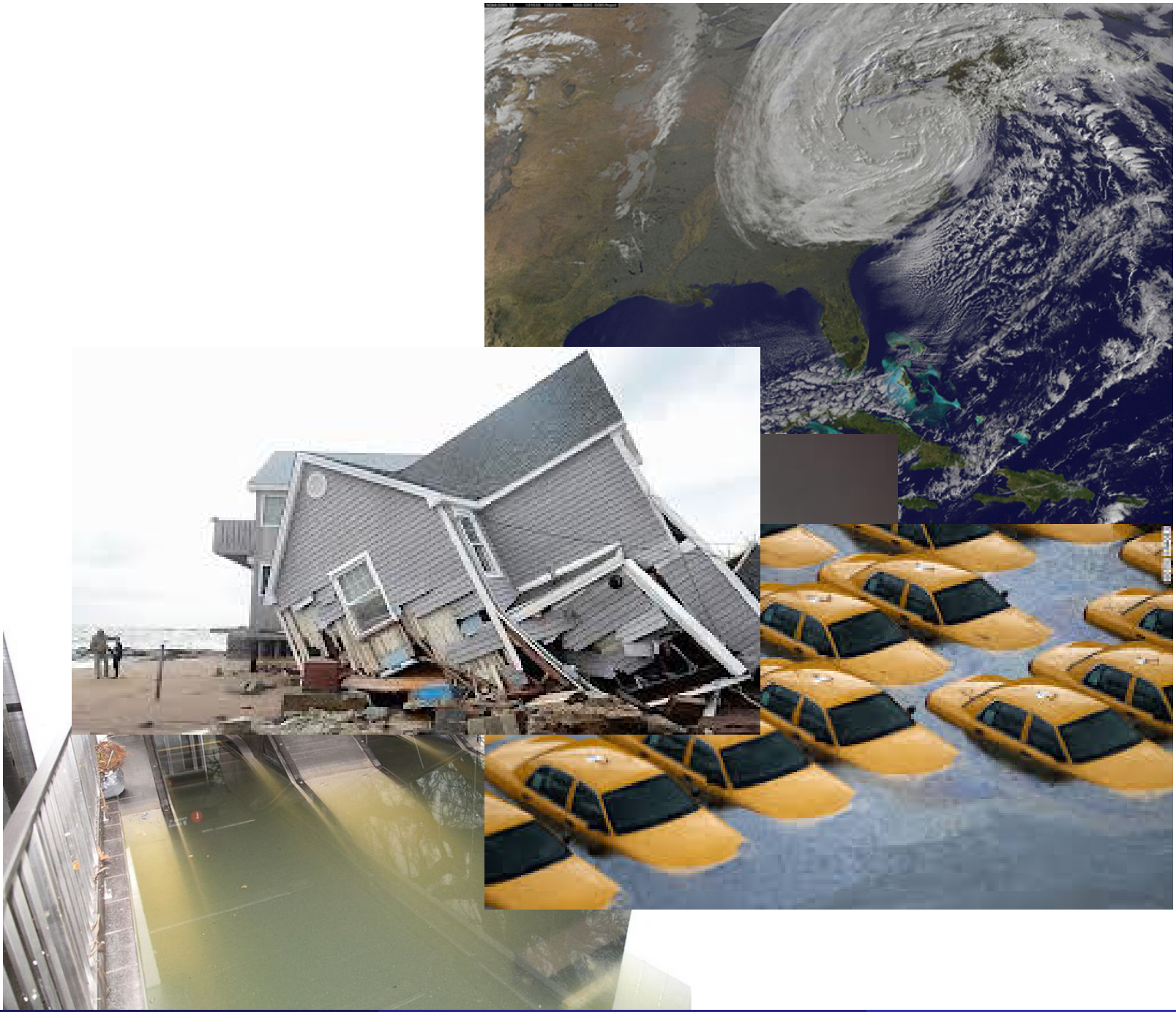


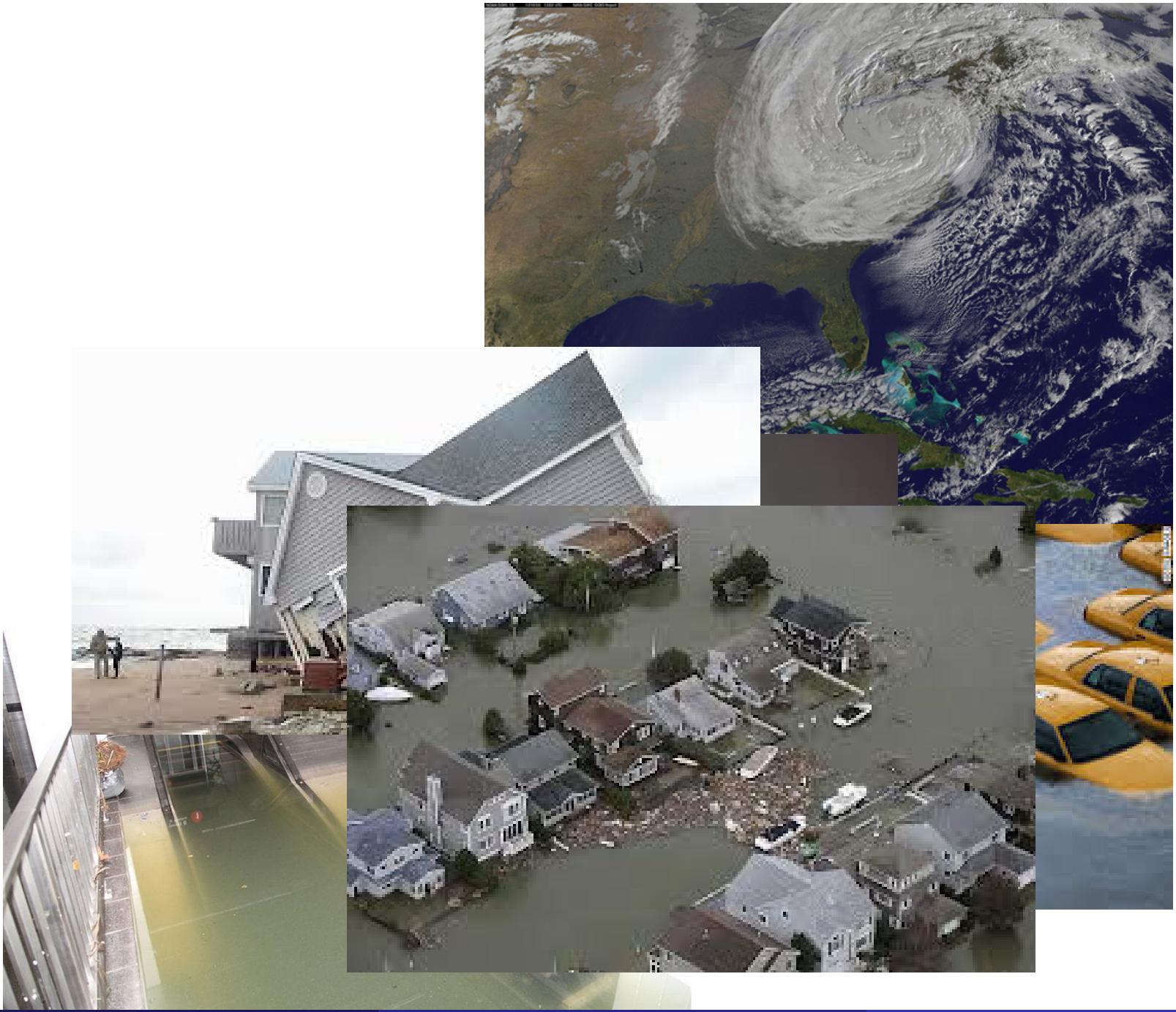

Uptown &
The Bronx
1






Uptown &
The Bronx
1





a tale of two (or three or four) cities

- business as usual (Manhattan above 39 St) vs. inconvenience vs. devastation
- need to identify (map) areas most hard hit in as close to real time as possible under ”austere” conditions
- housing damage
- power outages
- identify and clean data
- present in informative way

data

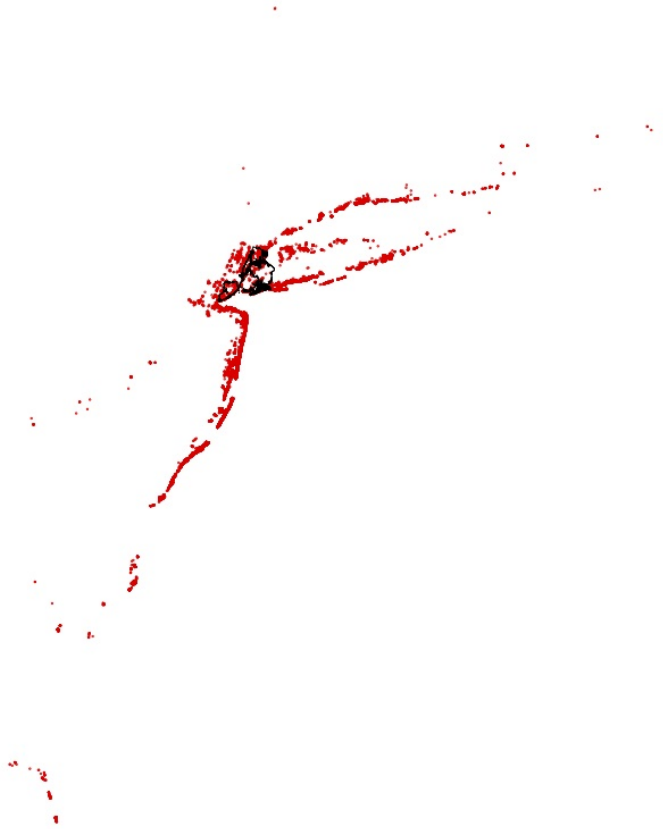
- New York City borough boundaries at Bytes of the Big Apple
- Federal Emergency Management Administration (FEMA) flyover surveillance data via Google Crisis Maps
  - getting FEMA's KML files directly into R problematic
    - failed approaches: read using OGR and GDAL (returned errors), convert to shape file in GRASS using v.out.ogr (didn't retain attribute data), parse KML file using xml package (would eventually have worked by didn't have time to learn the xml language elements necessary)
  - finally, located (with help from Jonathan Sury at National Center for Disaster Preparedness) google data directory of shape files
- 2 power outage data files from Consolidated Edison (Con Ed) converted from html
  - place-name coordinates, and number of customers, number of outages, place names

preparing the housing data

```
library(maptools)
library(rgdal)

fema.points<-readOGR("../femaPoints/", "femaPoints")
boros<-readOGR("../nybb/", "nybb")

fema.points<-spTransform(fema.points,
CRS("+proj=longlat +datum=NAD83"))
boros<-spTransform(boros,
CRS("+proj=longlat +datum=NAD83"))
```

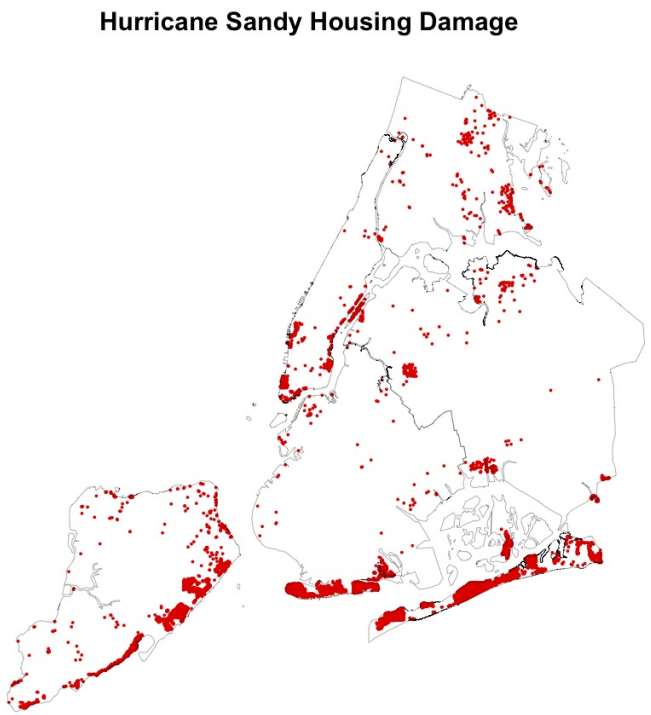
first plot

```
plot(fema.points,col="red", pch=20, cex=.1)
plot(boros, add=T, lty=1, lwd=.5)
```



a better first plot

```
#vignette("over")
plot(fema.points[boros,], col="red", pch=20, cex=.3)
plot(boros, add=T, lty=1, lwd=.2)
title(main="Hurricane Sandy Housing Damage",
sub="FEMA Flyover Data November 2012")
```



approach to a 2-D density map

- load ggplot2
- read in housing damage spatial data restricting to nyc
- convert to data frame (so will play with ggplot)
- build map layer by layer
- base map
- *fortify()* to convert nyc boro shapefiles to dataframe and overlay
- clean up a bit, and plot

code for 2-D density map

```
library(ggplot2)
fema.nyc<-fema.points[boros,]
fema.nyc.df<-as.data.frame(fema.nyc)
 #print(qplot(fema.nyc.df$LONGITUDE, fema.nyc.df$LATITUDE))
```

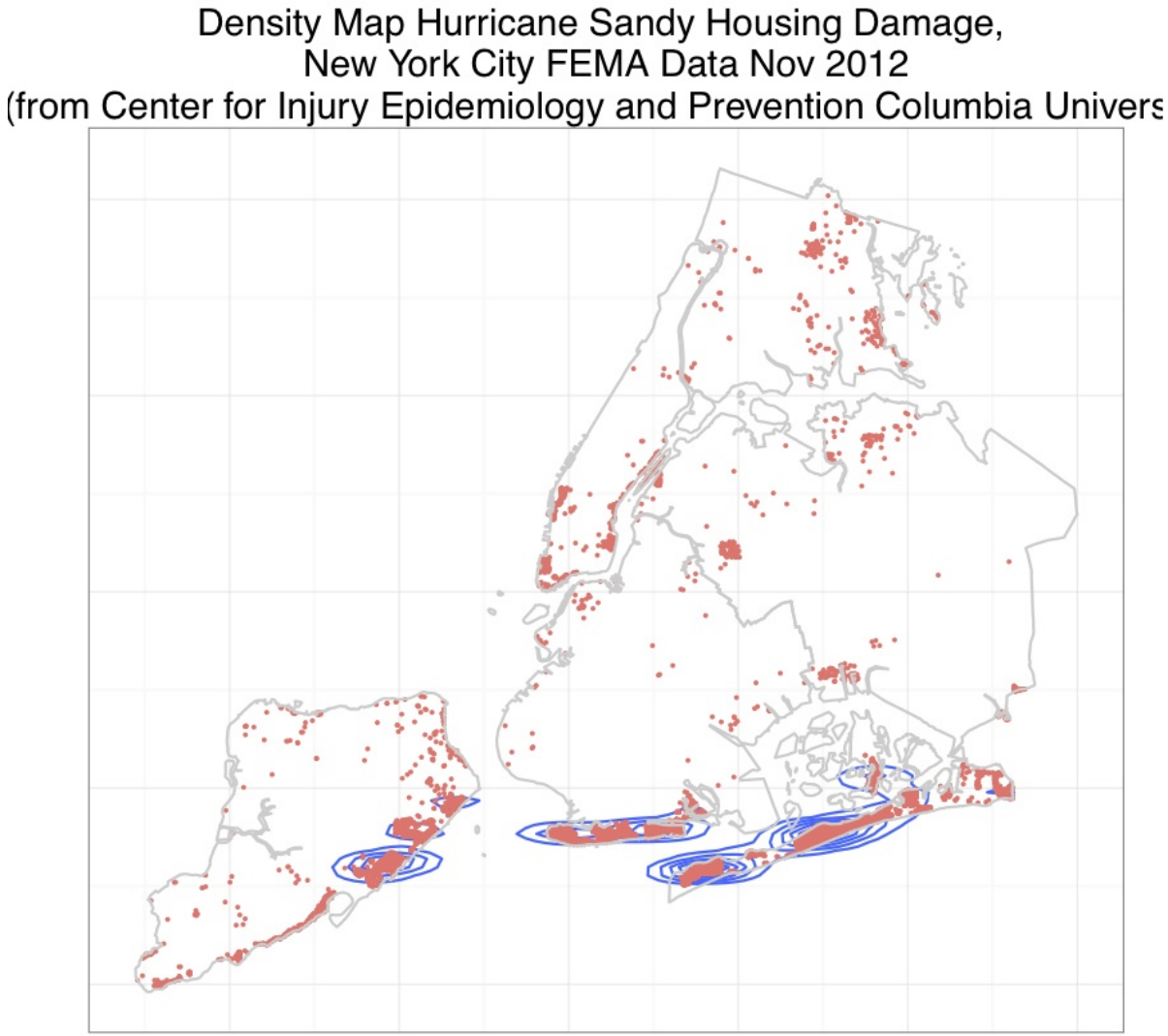
```
p<-ggplot() +
geom_density2d(aes(x=LONGITUDE, y=LATITUDE), data=fema.nyc.df) +
geom_point(aes(x=LONGITUDE, y=LATITUDE, col="red"), data=fema.nyc.df,
size=.9) +
theme(axis.text.x = element_blank(), axis.text.y = element_blank(),
axis.ticks = element_blank()) +
theme(panel.background = element_rect(colour = NA)) +
xlab("")+ylab("") +
theme(legend.position="none") +
ggtitle("Hurricane Sandy Housing Damage, New York City, Nov 2012")
  # print(p)
```

```
gpclibPermit()
boros_df <- fortify(boros, region='BoroCode')
```

```
p1<-p+geom_polygon(data=boros_df,aes(long, lat,group=group),fill="NA",
color="#CDCDCD")
 # print(p1)
```

```
p2<-p1+theme_bw()+
theme(axis.text.x = element_blank(), axis.text.y = element_blank(),
axis.ticks = element_blank()) +
theme(panel.background = element_rect(colour = NA)) +
xlab("")+ylab("") +
theme(legend.position="none") +
ggtitle("Density Map Hurricane Sandy Housing Damage,
\n New York City FEMA Data Nov 2012
\n (from Center for Injury Epidemiology and
Prevention Columbia University)" )
print(p2)
```

2-D density map of housing damage



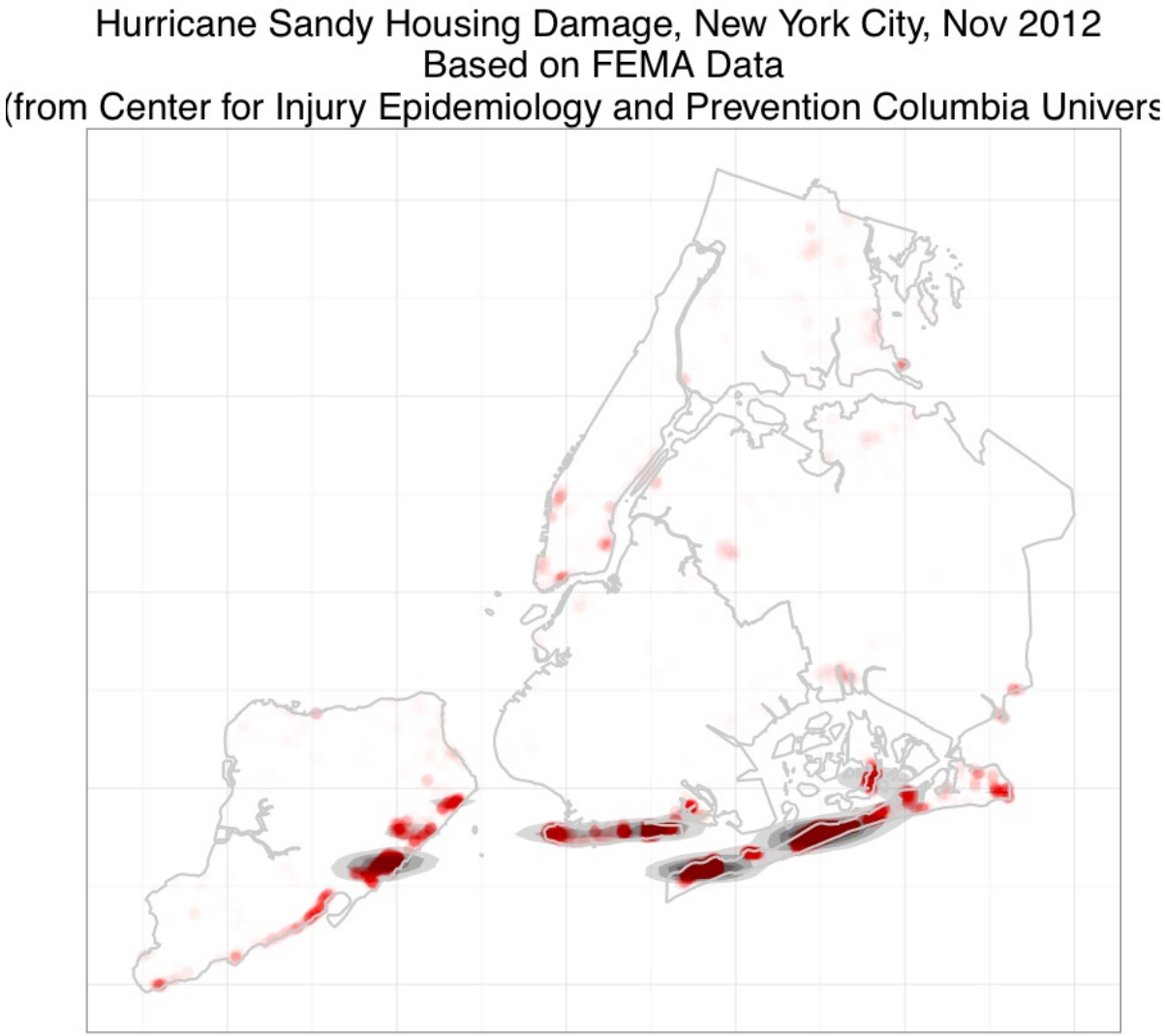
approach to ”weather-map” effect

- stolen shamelessly from David Kahle
  - see Jo fai Chow’s ggplot wrapper for crime maps...
- more intuitive (?less informative)
- again, build layer by layer
- establish base map with density plots
- smooth the levels of the density polygons using the alpha function
- overlay the points, color them red, and blur overlaid points with *alpha()*
- overlay borders, clean up, title

code for ”weather-map” effects

```
p<-ggplot(fema.nyc.df, aes(x=LONGITUDE, y=LATITUDE))

p1<-p+stat_density2d(aes(alpha=..level..), geom="polygon")
 # print(p1)

p2<-p1+scale_alpha_continuous(limits=c(0,500),breaks=seq(0,500,by=50))
 # print(p2)

p3<-p2+geom_point(colour="red",alpha=0.009)
 # print(p3)

p4<-p3+geom_polygon(data=boros_df,aes(long, lat,group=group),fill="NA",
color="#CDCDCD")
 # print(p4)

p5<-p4+theme_bw()+
theme(axis.text.x = element_blank(), axis.text.y = element_blank(),
axis.ticks = element_blank()) +
theme(panel.background = element_rect(colour = NA)) +
xlab("")+ylab("") +
theme(legend.position="none") +
ggtitle("Hurricane Sandy Housing Damage, New York City, Nov 2012
 \n Based on FEMA Data
\n (from Center for Injury Epidemiology and Prevention
Columbia University)" )
print(p5)
```

”weather-map” of Hurricane Sandy Housing Damage NYC



approach to mapping power outages

- read .csv files
- merge outage data file with centroids for place names, restricting to nyc
- convert to data frame for ggplot2
- variable of interest is proportion of customers without power 1 week after the storm
- first layer maps centroid coordinates to filled circles sized by proportion of customers without power
- overlay boro boundaries
  - NB: have to turn off ggplot feature of inheriting aesthetics from previous map (”inherit.aes=F”)
- print

code for power-outage bubble plot

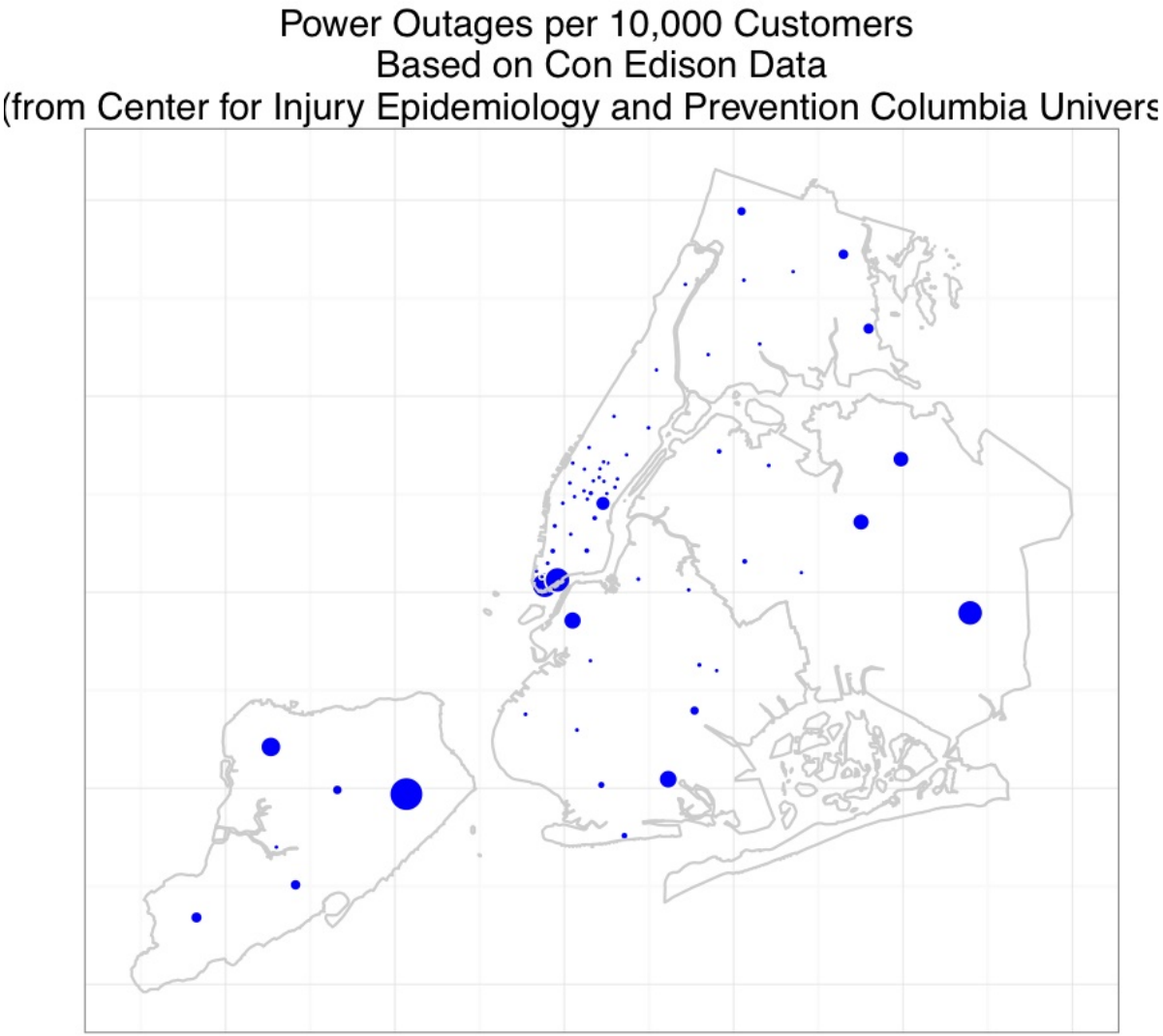
```
outages<-read.csv(".../conEdOutages.csv",header=T, stringsAsFactors=F)
places<-read.csv(".../conEdCoordinates.csv",header=T, stringsAsFactors=F)

conEd<-merge(outages, places, by="location", all.y=T)
conEd$outages[is.na(conEd$outages)]<-0
conEd$customers[is.na(conEd$customers)]<-.1
conEd$propOut<-conEd$outages/conEd$customers*100
summary(conEd$propOut)

coordinates(conEd)<-~long+lat
proj4string(conEd)<-CRS("+proj=longlat +datum=NAD83")

 # plot(conEd[boros,],col="blue", pch=20, cex=.9)

conEd.nyc<-conEd[boros,]

conEd.nyc.df<-as.data.frame(conEd.nyc)
```

```
p6<-ggplot(conEd.nyc.df, aes(x=long, y=lat, size=propOut, label=location,
 shape=21))+scale_shape_identity()
 # note need scale set to identity
p7<-p6+geom_point(color="white", fill="blue")

 # print(p7)
```

```
p8<-p7+geom_polygon(data=boros_df,aes(long, lat,group=group),fill="NA",
color="#CDCDCD", inherit.aes=F)
```

```
p9<-p8+theme_bw()+
theme(axis.text.x = element_blank(), axis.text.y = element_blank(),
axis.ticks = element_blank()) +
theme(panel.background = element_rect(colour = NA)) +
xlab("")+ylab("") +
theme(legend.position="none") +
ggtitle("Power Outages per 10,000 Customers
\n Based on Con Edison Data
\n (from Center for Injury Epidemiology and Prevention
Columbia University)" )
print(p9)
```

power outage map



R spatial data

- spatial data frame - fundamental spatial R object
  - translates Cartesian coordinates to geography
- 3 important attributes
  - x/y become lat/long
  - Coordinate Reference System (CRS) relates lat/long to earth
  - “bounding box” to display the data

R spatial packages

- “base” packages - sp, maptools, rgdal
  - useful utilities to read in, transform, display, manipulate spatial data
- spatial statistics - gstat, geoR, spBayes
- areal data - spdep, DCluster, ade4, SpatialEpi
- point pattern data - spatial, splancs. spatstat

SpatialPointsDataFrame object

- spatially-referenced points with attribute data
- sp:coordinates() will convert R dataframe to SpatialPointsDataFrame
- Spatial objects are *S4* objects
- note in str() output, uses the @ convention
- e.g. names(meuse@data)
- some special spatial tools, e.g select.spatial() to identify points
- generic plot() return spatially-referenced results, e.g plot()
- . . . but special spatial plots are available

```
spplot(meuse, "zinc")
bubble(meuse, "lead")
```

coordinate reference systems

- definitions for the location of a point on the earth
- *reference ellipsoid* - defines shape of the earth
- *datum* - reference point(s) on earth's surface
- unit of measurment, e.g. meters, miles
- projection - how spherical earth projected onto 2 dimensions (e.g. Mercator's)
- sp:proj4string() uses *rgdal* library to specify a CRS
- sp:spTransform() will change from one CRS to another
  - e.g.

```
proj4string(meuse) <- CRS("+init=epsg:28992")
```

conclusions

- just because you *can* map, doesn't mean you *should* map
- post-disaster surveillance a good place for maps
- it's all about the data
  - note no power outages hardest-hit areas (not serviced by Con Ed)
- if you do map, map in R, and consider ggplot as a tool
- if you map in R, buy Bivand, Prebesma and Gomez-Rubio

additional resources

- Roger Bivand - very highly recommended, and the basis for much of this material
- notes and tutorial
- chapter on spatial mapping of overdose deaths
- course exercise