Crowdsourcing public perceptions of space and safety

David Buil-Gil and Reka Solymosi
Department of Criminology, University of Manchester, UK

02/05/2020

Abstract

Digital platforms allow recording large volumes of data in relatively little time at a very small cost, and data generated through crowdsourcing is currently utilized for a variety of functions ranging from strategic police management to academic research. In criminological research, crowdsourcing projects have been deployed to harness people's experiences with crime and their perceptions about space and safety. Although crowdsourced data have some key strengths over data recorded by traditional survey methods (e.g., precise spatial data, reduced cost), their unique mode of production is also associated with certain limitations or weaknesses that, if uncontrolled, may affect the validity of such measures and the reliability of final results. This chapter reviews some published literature about the use of crowdsourcing for criminological research, discusses the main strenghts and limitations of data produced from crowdsourcing platforms, and presents a step-by-step exemplar study in `R` software using crowdsourced perceptions of safety in Atlanta, Georgia. Crowdsourcing offers many advantages over traditional survey methods to study perceptions of space and safety, but this exemplar study also shows how crowdsourced data may be affected by participation inequality, under-representation of areas and participation decrease. New methods are thus needed to account for and mitigate potential sources of bias in crowdsourced data.

Keywords: Fear of crime, perceived safety, crime mapping, open data, GIS, Atlanta

Full reference: Buil-Gil, D., & Solymosi, R. (2020). Crowdsourcing public perceptions of space and safety. In E. Groff & C. Haberman (Eds.), *The Study of crime and place: A methods handbook*. Temple University Press.

Contact details: David Buil-Gil. G18 Humanities Bridgeford Street Building, Cathie Marsh Institute for Social Research, University of Manchester. E-mail address: *david.builgil@manchester.ac.uk*

ORCID IDs: David Buil-Gil: 0000-0002-7549-6317. Reka Solymosi: 0000-0001-8689-1526.

1. Introduction

Crowdsourcing refers to the practise of enlisting the knowledge, experience or skills of a large number of citizens (*the 'crowd'*) to achieve a common goal or cumulative result, usually via a platform powered by online technologies, mobile phones, social media or a website (Howe 2006). Digital platforms allow recording large volumes of data in relatively little time at a very small cost, which explains why data generated through crowdsourcing is currently utilized for a variety of functions ranging from academic research to policy making and emergency management (Brabham 2008; Goodchild 2007; Hecker et al. 2019). As an example, during the 2007-2009 wildfires in the Santa Barbara area, California, residents shared their real-time knowledge about the location of fires and emergency shelters via various online forums and websites, which proved to be an invaluable source of information for disaster response (Goodchild and Glennon 2010). Similarly, crowdsourcing projects have been deployed to harness people's experiences with crime and their perceptions about space and safety (e.g., Solymosi and Bowers 2018; Williams, Burnap, and Sloan 2017). In this chapter we review some published literature about the use of crowdsourcing for criminological research, discuss the main strenghts and limitations of data produced from crowdsourcing platforms, and present a step-by-step

exemplar study in R software (R Core Team 2020) using crowdsourced perceptions of safety in Atlanta, Georgia.

In criminological research, crowdsourcing analysis has been primarily used to harness data about various forms of crime and antisocial behavior and to process information about citizens' perceptions and emotions about crime, thus allowing researchers to devise new explanations of crime and perceived safety. On one hand, open data recorded from social media and online forums enables detecting various forms of online crimes (e.g., hate speech towards minority groups; Miró-Llinares, Moneva, and Esteve 2018), and even associate patterns of online communication with offline serious offences (Bendler et al. 2014; Williams et al. 2020). Moreover, researchers have shown how users' online communication behaviors can be crowdsourced and analyzed for measuring the breakdown of social and physical order at detailed spatial scales (Erete et al. 2016; Williams, Burnap, and Sloan 2017). Criminologists can use large crowdsourced datasets to detect and explain criminal activity and disorder.

On the other hands, those researchers interested in the public perceptions about space and safety explore the use of volunteered crowdsourced information to explain the citizens' perceptions and emotions about crime. Criminologists have long known that public emotions about crime are not always explained by the prevalence and harms of crime, but instead fear of crime emotions are the result of individual predispositions to experience negative emotions about crime, which become fear of crime episodes under the presence of certain social and situational influences (Gabriel and Greve 2003; Hale 1996; Hough 2004). The public perceptions and emotions about crime have traditionally been analyzed by using surveys and interview-type qualitative approaches (see Gabriel and Greve 2003; Warr 2000), but these methods are costly and may be limited to capture the time and context-specific emotional reactions of fear of crime and the citizens' behavioural responses to such emotions of fear (e.g., avoidance behaviors, acquiring alarm systems or weapons). As an alternartive, some researchers have endorsed the use crowdsourcing and app-based tool to record data about the places and times in which episodes of fear of crime are more frequent, in order to fully conceptualize and operationalize the fear of crime as a function of individuals and their immediate environment (Solymosi and Bowers 2018; Solymosi et al. 2020).

To mention only a few examples of crowdsourcing platforms deployed to record data about perceptions and emotions about crime, Hamilton et al. (2011) developed a mobile phone app to record public perceptions of crime on public transportation in Melbourne, Australia. Similarly, Solymosi, Bowers, and Fujiyama (2015) designed an app and asked participants to report their worry about crime, which allowed authors to map the users' fear of crime across different areas of London, UK. Salesses, Schechtner, and Hidalgo (2013) designed and recorded data from the Place Pulse platform, which asks participants to choose 'which place looks safer' between two images taken from Google Street View. Then, Salesses, Schechtner, and Hidalgo (2013) produced a map of perceived safety in New York. In this chapter we will also use data recorded from Place Pulse to analyse perceived safety in Atlanta. Birenboim (2016) developed a mobile app to record data about the perceptions of security of attendees at a music festival in Jerusalem, Israel. Gómez et al. (2016) designed a collaborative web-based tool that allowed the citizens of Bogotá, Colombia, to report those areas in which they feel less safe. And Solymosi, Bowers, and Fujiyama (2017) analysed secondary data recorded from FixMyStreet, an online problem-reporting website, to examine perceptions of neighborhood disorder in London, UK. These are only a few examples, but there are many other crowdsourcing platforms that have been designed and utilized to study the fear of crime (see a review in Solymosi et al. 2020).

2. Strengths and weaknesses of crowdsourced data

Crowdsourced data about public perceptions of space and safety have some key strengths over data recorded from traditional survey methods (e.g., precise spatial data, information about immediate environmental variables, reduced cost). However, the mode of production of crowdsourcing is also associated with certain limitations or weaknesses that, if uncontrolled, may affect the validity of such measures and the reliability of final results (Buil-Gil, Solymosi, and Moretti 2020; Elliott and Valliant 2017). Amongst those limitations identified by researchers, the ones that are most commonly mentioned are related to the unequal participation arising from participants' self-selection, difficulty to interpret results, and under-representation of certain

areas and times (Solymosi et al. 2020). Others also identify that the number of participants in crowdsourcing projects tends to decrease over time (i.e., participation decrease), and some platforms have difficulties to engage participant and can only record small samples (e.g., Blom et al. 2010). We will briefly review some of these strengths and weaknesses of crowdsourcing projects and illustrate some of them with crowdsourced data about perceptions of safety in Atlanta.

Solymosi et al. (2020) conducted a systematic review of 27 studies utilizing or discussing the use of crowdsourcing to study perceptions and emotions about crime. The most frequent strength of crowdsourcing and app-based methods identified by researchers was that these techniques allow capturing the spatial-temporal specific nature of fear of crime, which was identified by 23 out of 27 papers. It is well known today that 'mental events' of fear of crime are qualitatively and quantitatively different from 'mental states' of worry or anxiety: whereas emotional reactions of fear take place due to the presence of situational and context-specific elements that make the person feel threatened (Castro-Toledo et al. 2017), anxieties and worries about crime refer to more general concerns about immediate and future risks (Buil-Gil et al. 2019; Hough 2004). Crowdsourcing techniques powered by digital technologies enable recording emotions of fear of crime in their precise space and time, which ultimately allows to "un-erroneously associate them [experiences of fear] with elements of the environmental backcloth such as incivilities, crime, and disorder" (Solymosi, Bowers, and Fujiyama 2015, 198). This is related to another important strength of crowdsourcing: these techniques allow recording data about the architechtural features and environmental characteristics of spaces where mental events of fear are more common (Chataway et al. 2017; Traunmueller, Marshall, and Capra 2015).

Another obvious advantage of crowdsourcing is its reduced cost of data collection. Large volumes of data can be recorded at a very low cost. Dubey et al. (2016), for example, analyzed more than 350,000 votes of perceived safety recorded from the Place Pulse platform; and Solymosi, Bowers, and Fujiyama (2017) analyzed more than 275,000 reports of disorder in London. These large samples are very costly to record by using traditional probability surveys. In this chapter we will illustrate how to download data about more than 1.5 million votes registered from the Place Pulse platform, and we will analyze more than 35,000 votes of perceived safety in Atlanta.

Crowdsourcing offers advantages over other methods to record data about perceptions of space and crime, but its unique mode of production also creates certain challenges that may affect the validity and precision of recorded data. Perhaps the main weakness of data recorded from crowdsourcing is related to participants' self-selection. Probability surveys are carefully designed to select participants randomly, which means that all units in the population have equal probabilities of being chosen; whereas crowdsourcing projects harness data from non-probability samples who decide when and where to share their perceptions and emotions, and whether they want to participate at all (Elliott and Valliant 2017). The mode of production of crowdsourced data increases the risk of self-selection bias, and as a consequence males and young citizens tend to be overrepresented in these data (Chataway et al. 2017), and citizens from deprived areas are generally less represented than persons from wealthy neighborhoods (Solymosi and Bowers 2018). For example, Salesses, Schechtner, and Hidalgo (2013) observed that 78.3% of participants who informed about their gender when using the Place Pulse platform were males, and Solymosi, Bowers, and Fujiyama (2017) highlight that only 26.0% of those who informed about their gender when reporting instances of disorder via FixMyStreet were females.

Those crowdsourcing platforms that allow participants to vote multiple times may also suffer from participation inequality (or unequal participation). In order words, "it is often observed that few users are responsible for most crowdsourced information, while the majority participate only a few times" (Buil-Gil, Solymosi, and Moretti 2020, 6). To illustrate this, Dubey et al. (2016) show that 6,118 of the 81,730 persons who used the Place Pulse platform participated only once, while 30 users participated more than 1,000 times and the most prolific user voted 7,168 times. Solymosi, Bowers, and Fujiyama (2017) also show that one fourth of all FixMyStreet reports are produced by one percent of participants, and 73% of participants contribute only once.

The fact that users of crowdsourcing projects can decide where and when to participate also explains why certain areas and times are underrepresented. For instance, it is likely that app-based platforms fail to capture data from high-crime-density areas, since participants may avoid those places where they feel more exposed

to crime (Innes 2015). The routine activities of participants are also reflected on an under-representation of votes at night (Blom et al. 2010).

Finally, there are ethical considerations that may arise from the use of crowdsourcing, which are related to the user's privacy (e.g., risk of participants' identification) but also concerns that these techniques may sensitize participant and increase their fear of crime by asking them to constantly think about crime-related risks (Jackson and Gouseti 2015; Solymosi et al. 2020).

3. Crowdsourcing perceptions of safety: Step-by-step example in R

In order to illustrate the use of crowdsourcing in criminological research, we present an exemplar study using data recorded by the Place Pulse 2.0 platform (Salesses, Schechtner, and Hidalgo 2013). This section will introduce the Place Pulse project and provide R codes to download, explore and clean this source of crowdsourced data. Then, we will analyse the spatial distribution of crowdsourced perceptions of space and safety in Atlanta and illustrate with examples some of the known issues of crowdsourced data (i.e., participantion inequality, under-representation of certain areas and participation decrease).

3.1 The Place Pulse project

Place Pulse 2.0 was an online crowdsourcing platform designed to record data about citizens' perceptions of safety, beauty, wealth, liveability, boredom and depression in urban areas. Two images were shown to participants, who then were asked to answer *'Which place looks safer?'* (see Figure 1). Participants could also be asked which of the two images looked wealthier, more beautiful, more boring, livelier or more depressing, but we will focus on perceptions of space and safety in this chapter. Images were selected randomly from Google Street View across 56 cities from 28 countries, and these captions were originally taken between 2007 and 2012. The platform recorded all votes in a public dataset, but respondents did not provide any further information about themselves, which means that we do not know the social and demographic characteristics of participants. Place Pulse used to be hosted in an open website (http://pulse.media.mit.edu/) and anyone could use it, but the platform was closed in late 2019. We have been granted access to all the data recorded between May 28th 2013 and August 22nd 2019 to write this chapter. All data have also been uploaded onto an open repository with consent of the data producers (Salesses, Schechtner, and Hidalgo 2013).

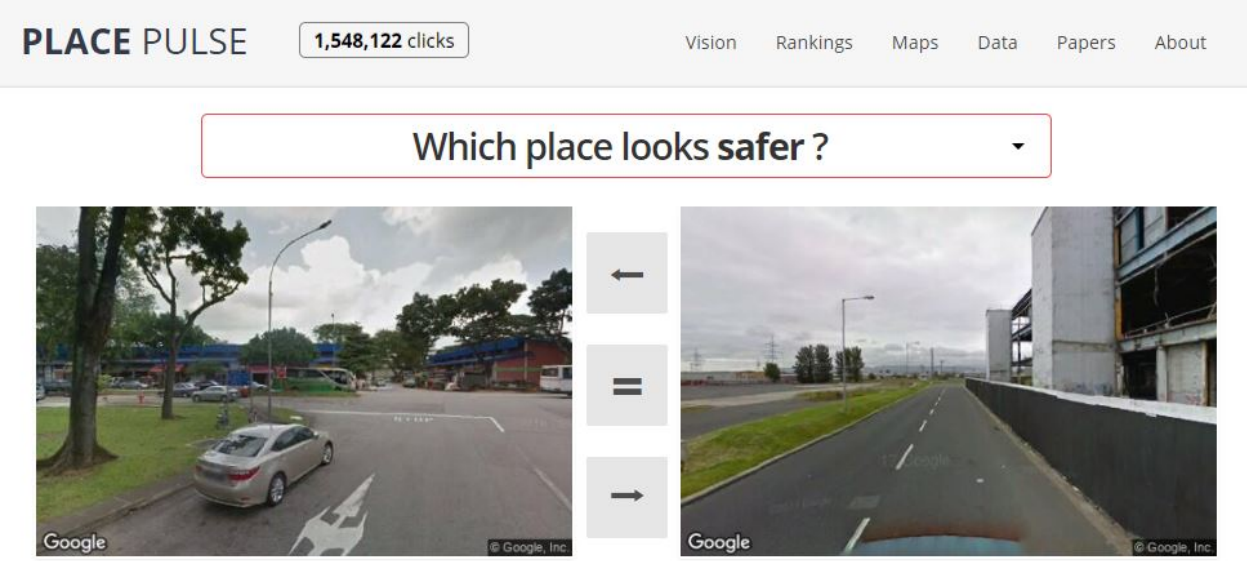


Figure 1: Figure 1: Place Pulse website

3.2 Download and explore Place Pulse data

We have saved all Place Pulse data (more than 1.5 million votes) in a data repository on FigShare. You can download this directly into `R` by using the `read.csv()` function. It is a large file so it may take some minutes to read in.

```
pp_data <- read.csv('https://ndownloader.figshare.com/files/21739137')
```

This dataset includes 17 variables, but we will only use some of them. A unique identification code was given to each participant (*'voter_uniqueid'*) and image (*'place_id_left'* for images in the left side of the pairwise comparison and *'place_id_right'* for images in the right part). The columns *'place_name_left'* and *'place_name_right'* specify the city of each photograph. The column *'choice'* shows if the user perceived the image in the left or right to be 'safer' (participants could also answer 'equal'), and the column *'study_question'* allows studying perceptions about different variables (e.g., safety, wealth, beauty). The columns *'day'* and *'time'* specify the moment when each vote took place, and the columns *'long_right'*, *'lat_right'*, *'long_left'* and *'lat_left'* show the longitude and latitude of both photographs.

We can begin by examining which cities were more frequenty assessed within the Place Pulse platform. This can be checked for images in the left side of the pairwise comparison first, and then for images in the right. We can use the `group_by()` and `summarize()` functions from `dplyr` package (Wickham, François, et al. 2020) to check this:

```
pp_data %>%
  group_by(place_name_left) %>% # categories based on cities on the left
  summarize(Count = n()) %>% # count number of units in each category
  top_n(3) # print 3 most frequent categories
```

```
## # A tibble: 3 x 2
##   place_name_left Count
##   <fct>           <int>
## 1 Atlanta         56992
## 2 Berlin          55265
## 3 Tokyo           53817
```

Atlanta was the city with the largest number of votes amongst those images that appeared in the left part of the comparison, but we can also check this for those images shown in the right side:

```
pp_data %>%
  group_by(place_name_right) %>% # categories based on cities on the right
  summarize(Count = n()) %>% # count number of units in each category
  top_n(3) # print 3 most frequent categories
```

```
## # A tibble: 3 x 2
##   place_name_right Count
##   <fct>            <int>
## 1 Atlanta          57140
## 2 Berlin           55583
## 3 Tokyo            53514
```

Atlanta was indeed the city with the largest number of votes. We can also check which variables (e.g., safety, beauty, wealth) were more frequently assessed by participants:

```r
pp_data %>%
  group_by(study_question) %>% # categories based on study questions
  summarize(Count = n()) %>% # count number of units in each category
  arrange(desc(Count)) # print in descending order
```

```
## # A tibble: 7 x 2
##   study_question   Count
##   <fct>            <int>
## 1 safer           509961
## 2 livelier        366802
## 3 more beautiful  220604
## 4 wealthier       174758
## 5 more depressing 149355
## 6 more boring     144060
## 7 <NA>               183
```

We see that safety was the most commonly assessed variable, with 509,961 votes in total. In this chapter we will examine reports of safety in the city of Atlanta. Before analysing the data, however, we can also examine if participants were more inclined to vote for images in the left or right part of the platform. In other words, we analyse if responses were biased by the position in which images were shown on the website.

```r
pp_data %>%
  group_by(choice) %>% # categories based on vote (right, left or equal)
  summarize(Count = n()) %>% # count number of units in each category
  top_n(3) # print 3 most frequent categories
```

```
## # A tibble: 3 x 2
##   choice  Count
##   <fct>   <int>
## 1 equal  206147
## 2 left   668680
## 3 right  690792
```

The frequency of votes for left and right options is very similar, and thus we can conclude that the position of the image on the website platform does not appear to have an affect on participants' votes.

3.3 Cleaning Place Pulse data

When it comes to crowdsourced data, the first step is to clean the data to make it as complete and useful as possible to conduct our research projects. In other words, we want to prepare the data to allow us answering our research questions. For example, in this case, we want to map the perceived safety of areas in Atlanta. To do this, first we have to select the votes about safety for Atlanta.

Atlanta is the capital city of the State of Georgia, United States. In 2018, its estimated papulation was close to 500,000 residents, and it is the 37th most populated city in the United States. There are two main reasons why we are conducting this exemplar study in Atlanta: first, as shown above, it was the city with the largest number of votes in the Place Pulse platform; and second, there is available open data about the social, demographic and crime characteristics of Atlanta neighborhoods, which can be easily accessed to explain the patterns observed in Place Pulse data. Moreover, various papers have analysed the predictors of crime and fear of crime in this city, which can be used to interpret our results (see McNulty and Holloway 2000; Tester et al. 2011)

We can use the function `filter()` from `dplyr` to create a new dataframe that includes those pairwise comparisons in which either the image of the right or the image of the left, or both, are from Atlanta. Thus, we remove all those votes for images of other cities.

```
# select cases in which the image of the right or left is from Atlanta
pp_atl <- pp_data %>%
  filter(place_name_right == "Atlanta" | place_name_left == "Atlanta")
```

We will also focus on the ratings of areas as 'safer', as we are interested in people's perceptions of place and safety:

```
pp_atl_s <- pp_atl %>%
  filter(study_question == "safer") # select votes of 'safer'
```

You can see now that we have a dataframe of 37214 votes about the safety of places in Atlanta.

We are interested in analyzing the proportion of 'safer' votes in each neighborhood of Atlanta. In order to assign photographs to their neighborhood, we need to create two new columns that specify the longitude and latitude of each image of Atlanta being assessed. We will also create a new column that details whether each participant voted that the image of Atlanta was 'safer' or 'not safer' (i.e., less safe or equal) than the other photograph. Some pairwise comparisons, however, assessed two different images from Atlanta, which means that we will need to duplicate those votes to account for both images on the right and left of the pairwise comparison. First, we want to know the number of comparisons in which both images are from Atlanta. We can use the function `filter()` from `dplyr`, which we have also used above.

```
# create dataset in which both images are from Atlanta
pp_atl_s_dup  <- pp_atl_s %>%
    filter(place_name_right == "Atlanta" & place_name_left == "Atlanta")

# print the count of votes as a result
pp_atl_s_dup %>%
    summarize(count = n())
```

```
##   count
## 1   678
```

In total, 678 pairwise comparisons are based on two images from Atlanta, whereas 36,536 compare one image from Atlanta against a photogrpah from any other city. We will duplicate those comparisons in which both images were taken in Atlanta and attach them to two new datasets (one to assess the images on the right, and the other to rate the images on the left). For now, we also delete these duplicated cases from the main dataframe by using the `anti_join()` function from `dplyr`, which searches those units that exist in the newly created dataset `pp_atl_dup` and removes them from the main dataset of votes. We will merge all units together once all data have been cleaned.

```
# duplicate the new dataset
pp_atl_s_dup2 <- pp_atl_s_dup

# delete duplicated votes from main dataset
pp_atl_s <- pp_atl_s %>%
  anti_join(x = pp_atl_s, y = pp_atl_s_dup, by = "X")
```

Now, the main dataset includes only those pairwise comparisons in which only one image is from Atlanta. We create two new columns that specify the coordinates of the photograph from Atlanta. We use the `mutate()`

function from `dplyr` to create the new columns and the `if_else()` function to copy the coordinates of the left (or right) image depending on whether the left photoraph is from Atlanta or not.

```
# copy coordinates from left image if it is from Atlanta, otherwise copy from right image
pp_atl_s <- pp_atl_s %>%
    mutate(long_Atl = if_else(place_name_left == "Atlanta", long_left, long_right),
           lat_Atl  = if_else(place_name_left == "Atlanta", lat_left , lat_right))
```

Remember that we had previously created two new datasets with those votes in which both images are from Atlanta. The first dataset (`pp_atl_s_dup`) is used to assess the images in the left, while the second dataset (`pp_atl_s_dup2`) refers to the images in the right. We can then allocate the coordinates of the left image to two new columns in the first dataset of votes in which both images are from Atlanta:

```
# copy coordinates from left image
pp_atl_s_dup <- pp_atl_s_dup %>%
  mutate(long_Atl = long_left,
         lat_Atl  = lat_left)
```

And then we allocate the coordinates of the right image to the second dataset of pairwise comparisons between Atlanta pictures:

```
# copy coordinates from right image
pp_atl_s_dup2 <- pp_atl_s_dup2 %>%
  mutate(long_Atl = long_right,
         lat_Atl  = lat_right)
```

Before merging all the data into a single dataset, we will also create a new column that distinguishes those images of Atlanta that were assessed as 'safer' from those reported as 'less safe' or 'equal'. Later we will use this column to compute the proportion of 'safer' votes in each area. We do this first in the main dataset (for now, it only includes votes with one picture from Atlanta), by checking if the choice of each vote (i.e., 'left', 'right', or 'equal') corresponds to the position of the image from Atlanta. We assign a 1 if the user chose the image of Atlanta as 'safer', whereas a 0 is assigned when the image of a different city was chosen to be 'safer' and when users voted 'equal'. We also use the `mutate()` and `if_else()` functions from `dplyr`.

```
# if left image is from Atlanta and left image chosen as 'safer' or
# right image is from Atlanta and right image chosen as 'safer', assign 1, otherwise 0
pp_atl_s <- pp_atl_s %>%
  mutate(win = if_else((place_name_left  == "Atlanta" & choice == "left") |
                       (place_name_right == "Atlanta" & choice == "right"), 1, 0))
```

Similarly, in the `pp_atl_s_dup` dataset, we assign a 1 to those cases in which participants voted for the left image as 'safer' and a 0 otherwise, given that this dataset had been previously created to assess the left images in those pairwise comparisons in which both images are from Atlanta. And we do the same with the second dataset (`pp_atl_s_dup2`) that compares two images from Atlanta, which in this case refers to the image in the right.

```
# if participant voted for left image as 'safer', assign 1, otherwise 0
pp_atl_s_dup <- pp_atl_s_dup %>%
  mutate(win = if_else(choice == "left", 1, 0))

# if participant voted for righ image as 'safer', assign 1, otherwise 0
pp_atl_s_dup2 <- pp_atl_s_dup %>%
  mutate(win = if_else(choice == "right", 1, 0))
```

Now that our dataset has been cleaned and is ready to be analyzed, we can merge all the data together with the `rbind()` function.

```
pp_atl_s <- rbind(pp_atl_s, pp_atl_s_dup, pp_atl_s_dup2)
```

We have a dataframe of 37892 votes about the safety in Atlanta that is ready to be analyzed.

3.4 Map Place Pulse data

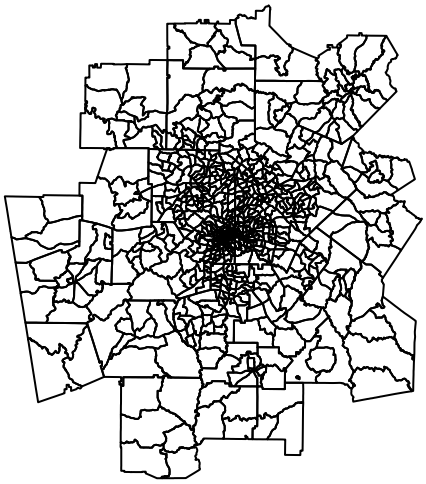
It is possible to use mapping techniques learned in other chapters (LINK WITH MAPPING CHAPTER) to map crowdsourced data (see also Solymosi, Bowers, and Fujiyama 2015). We will be using the `sf` (Pebesma 2020) and `ggplot2` (Wickham, Chang, et al. 2020) libraries in order to create a map of perceived safety of built environment across the areas of Atlanta.

First, acquire a shapefile for Atlanta. We can download the Atlanta Region census tracts shapefile from various sources. The Georgia Association of Regional Commission, for instance, publishes spatial data for Atlanta Region at the different spatial scales. You can go on their website to find out more about this boundary data: https://opendata.atlantaregional.com/datasets/census-2000-tracts-atlanta-region. We can download the shapefile directly using their Application Programme Interfact (or API) and the `geojson_sp()` function from `geojsonio` package (Chamberlain and Teucher 2020) - this is something discussed in greater detail on the chapter on Open Data (CHAPTER REF LANGTON AND SOLYMOSI, 2020). In case this API call fails (e.g., because of a network error), we have uploaded the original shapefile onto a Github repository: https://github.com/maczokni/crowdsourcing_pp_chapter/blob/master/geojson/Census_2000_Tracts_Atlanta_Region.geojson. And you can also download this file from the Worldmap platform of Harvard University: https://worldmap.harvard.edu/data/geonode:Atlanta_Census_Tracts_SHL using the the `st_read()` function of `sf` package. You can just use the code below:

```
library(geojsonio) # load 'geojsonio' package

# download geojson from Georgia Association of Regional Commissions open data
atl <- geojson_sp("https://opendata.arcgis.com/datasets/04b79404794f43959cda4f8c3f1817e6_49.geojson")

# download geojson from Github
#atl <- geojson_sp("https://github.com/maczokni/crowdsourcing_pp_chapter/raw/master/geojson/Census_2000_

# convert sp file into st
atl <- st_as_sf(atl)

# download json from WorldMap of Harvard University
#atl <- st_read("http://worldmap.harvard.edu/download/wfs/1824/json?outputFormat=json&service=WFS&reque
```

We can see what this file looks like by using the `plot()` function to plot the geometry of the 'atl' object we created, called with the `st_geometry()` function:

```
plot(st_geometry(atl),
     main = "Atlanta Region census tracts")
```



Now, to be able to plot the safety votes on this map, we first need to make our votes a spatial object, by specfying that the *'long_Atl'* and *'lat_Atl'* columns contain our longitude and latitude information. We use the `st_as_sf()` function for this:

```
points_atl_s <- st_as_sf(pp_atl_s, coords = c("lat_Atl", "long_Atl")) #geocode votes
```

In order to plot both these spatial layers (i.e., votes recorded from Place Pulse and Atlanta census tracts) on the same map, their coordinate reference systems (CRS) need to match. We can check these with the `st_crs()` function:

```
st_crs(points_atl_s) == st_crs(atl) #check if CRS is the same in both layers
```

```
## [1] FALSE
```

The function tells us that it is 'false' that both CRS are equal, but we can change this with the following line of code:

```
st_crs(points_atl_s) <- st_crs(atl)
```

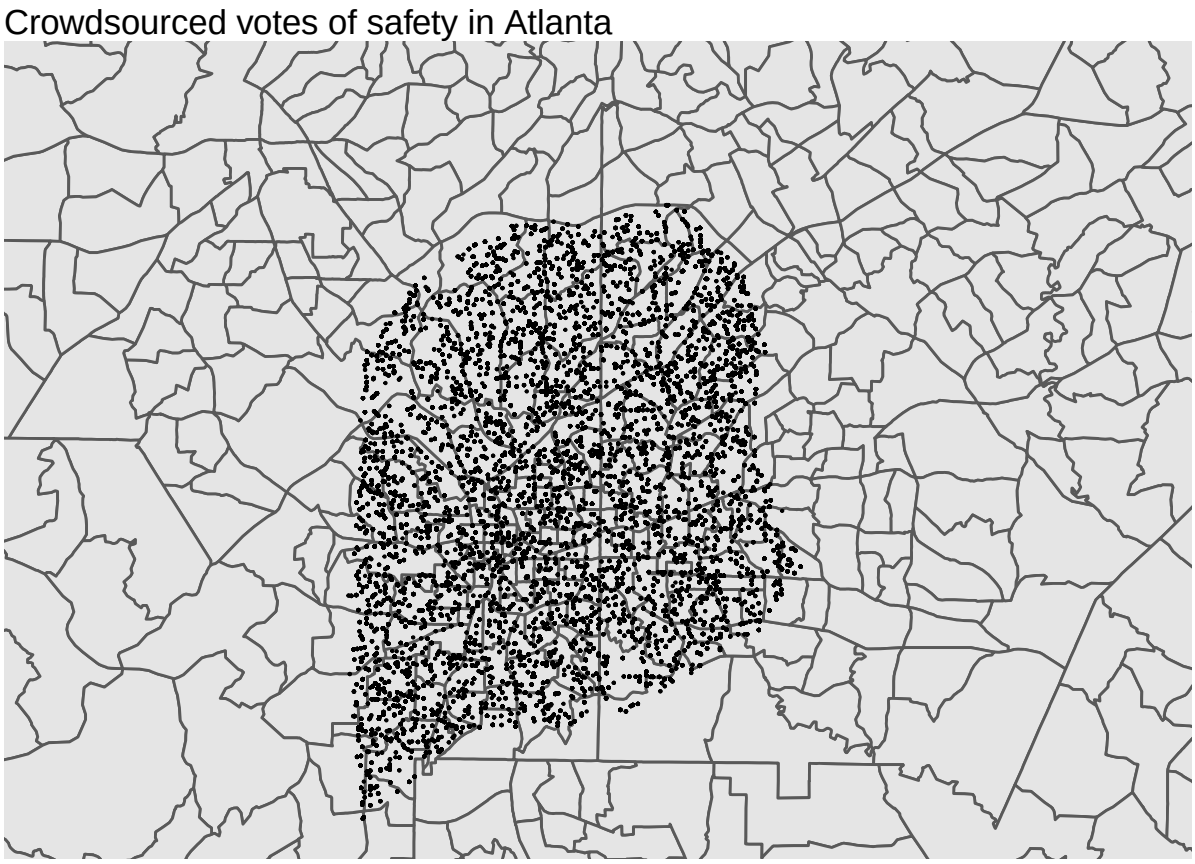
Now, if we check, they should have the same CRS:

```
st_crs(points_atl_s) == st_crs(atl) #check if CRS is the same in both layers
```

```
## [1] TRUE
```

Thus, we can now map our data. See (LINK WITH MAPPING CHAPTER) for more information about crime mapping.

```
map <- ggplot(data = atl) + geom_sf() + theme_void() +
       coord_sf(xlim = c(-84.7, -84), ylim = c(33.6, 34),
                expand = FALSE) #create map

map + ggtitle("Crowdsourced votes of safety in Atlanta") +
  geom_point(data = pp_atl_s, aes(x = lat_Atl, y = long_Atl),
                 size = .1) #plot map with points
```



This is a very busy map. Maybe instead we want to get some sort of average score for each census tract. We then calculate the proportion of 'safer' responses in each area. Note that Salesses, Schechtner, and Hidalgo (2013) suggest computing a Q-score per image corrected by the "win" and "loss" ratio of all photographs with which it is compared, but for the purpose of this chapter we will compute a simple proportion that will allow us to directly analyse the geographical distribution of perceived safety (see Buil-Gil, Solymosi, and Moretti 2020).

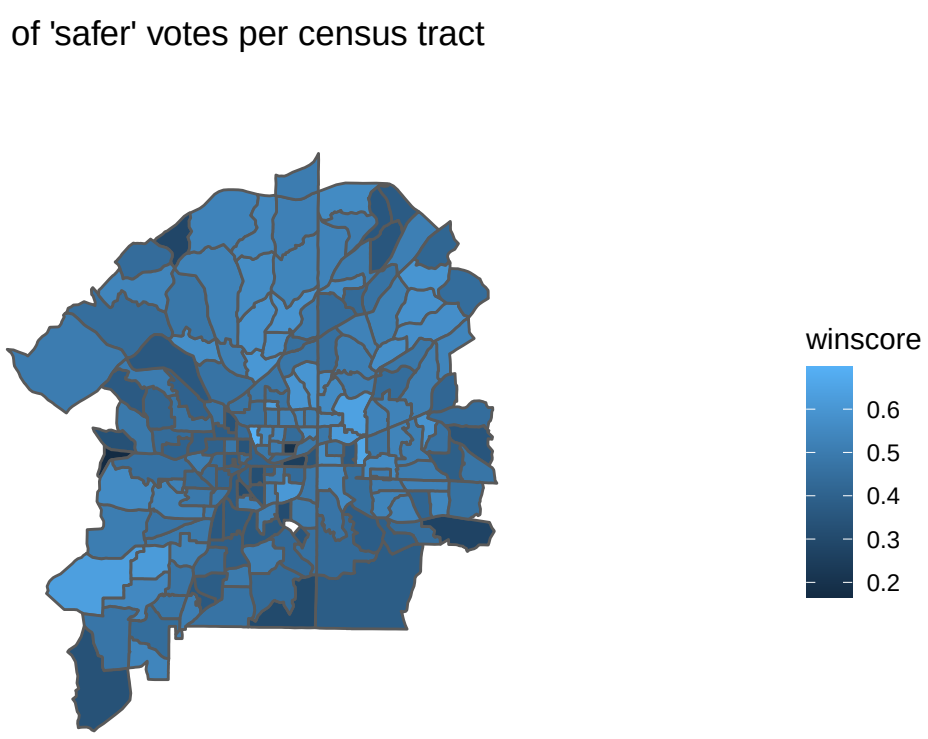
```
points_atl_s_nhood <- st_intersection(atl, points_atl_s) %>% # intersection of points and areas
                      group_by(TRACT) %>% # make groups based on tracts
                      summarise(winscore = mean(win, na.rm = TRUE), # proportion 'safer'
                                num_votes = n()) # count number of votes
```

We can add the average score of 'safer' reponses per area to the original shapefile of census tracts using the `left_join()` function of `dplyr`. We will also delete those census tracts in which we do not have any vote of perceived safety by using the `filter()` function.

```
# @REKA: CAN YOU ADD A SHORT NOTE HERE?
st_geometry(points_atl_s_nhood) <- NULL

# merge census tracts and Place Pulse votes based on 'TRACT' column
atl_pp_wins <- left_join(atl, points_atl_s_nhood, by = c("TRACT" = "TRACT")) %>%
  filter(!is.na(winscore)) # delete census tracts with 0 votes (NAs)
```

Finally, we can plot the propotion of 'safer' votes in each census tract.

```
ggplot(data = atl_pp_wins) +
  ggtitle("Proportion of 'safer' votes per census tract") +
  geom_sf(aes(fill = winscore)) +
  coord_sf(xlim = c(-84.7, -84), ylim = c(33.5, 34), expand = FALSE) +
  theme_void()
```

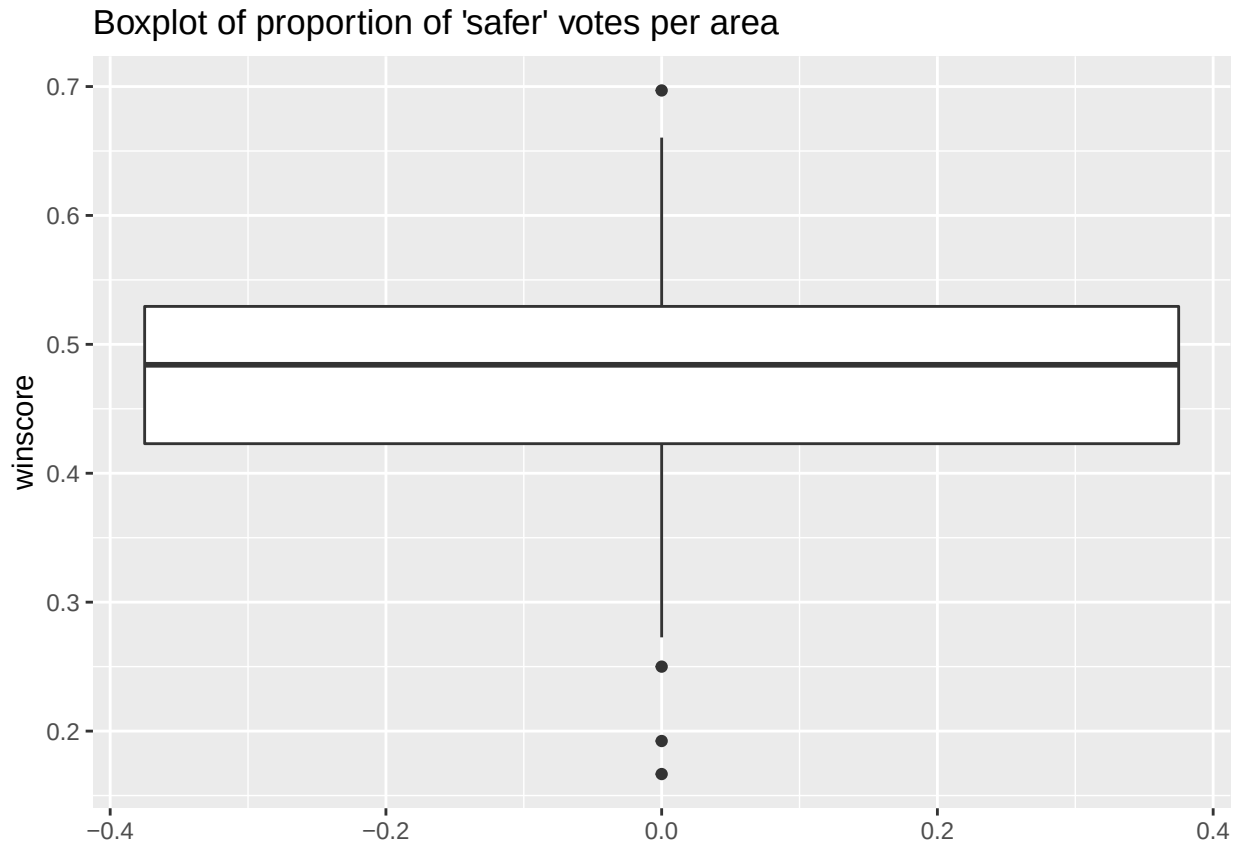


There are many more things one can do with these data. For example, we could look at the descriptive statistics using the `summary()` function, and produce a boxplot of the proportion of 'safer' votes in each area using the `geom_boxplot()` function from `ggplot2`.

```
summary(atl_pp_wins$winscore) # descriptive statistics: proportion 'safer' votes per tract
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.1667  0.4229  0.4841  0.4742  0.5294  0.6970
```

```r
p <- ggplot(atl_pp_wins, aes(y = winscore))

p + geom_boxplot() + ggtitle("Boxplot of proportion of 'safer' votes per area") # boxplot
```



Researchers may also be interested in analyzing the environmental characteristics of those places assessed by Place Pulse users to be the safest or the least safe. We show below some lines of code to print the coordinates of places rated as the safest and the least safe. To illustrate this, we only consider those places that have been evaluated by at least 20 participants, to increase the validity of these rates, but one may decide to reduce or increase this threshold number according to different criteria. For example, if we were interested in obtaining the coordinates of the five places rated as the safest we would need to change the last line of code by `head(5)`, or `tail(5)` to obtain the coordinates of the five least safe places.

```r
points_atl_s %>%
  mutate(geom_text = as.character(st_geometry(geometry))) %>% # spatial data into character
  group_by(geom_text) %>% # group votes based on spatial data
  summarize(winscore = mean(win, na.rm = TRUE), # calculate proportion of 'safer'
            num_votes = n()) %>% # count number of votes
  filter(num_votes >= 20) %>% # filter out pictures with less than 20 votes
  arrange(desc(winscore)) %>% # order by proportion of 'safer' responses
  filter(row_number() == 1 | row_number() == n()) # print lowest and highest score
```

```
## Simple feature collection with 2 features and 3 fields
## geometry type:  POINT
## dimension:      XY
## bbox:           xmin: -84.37582 ymin: 33.71869 xmax: -84.37231 ymax: 33.73143
```

```
## CRS:            unknown
## # A tibble: 2 x 4
##   geom_text                winscore num_votes            geometry
## * <chr>                       <dbl>     <int>          <POINT [°]>
## 1 c(-84.375818, 33.731433)    0.857        21 (-84.37582 33.73143)
## 2 c(-84.372314, 33.71869)     0.136        22 (-84.37231 33.71869)
```

This may be used by criminologists to observe the characteristics of those places assessed by Place Pulse participants as more or less safe. For example, we can use Google Street View [https://www.google.com/maps] to observe places rated as the least safe or safest amongst those rated at least 20 times (see Figure 2). In this case, the least safe place is characterized by very dense vegetation (which may be perceived to offer concealment for possible criminals and obstructs the view onto the ground), lack of exit routes to escape from potential threats, and an abandoned house with signs of physical disorder (see Fisher and Nasar 1992); whereas the safest place is a wider street of a residential area with direct visual access to most places around it (large prospect), lack of places for concealment of offenders and natural surveillance from houses (Welsh and Farrington 2004). We can do much more, but here we will focus on the specific issues to explore due to the crowdsourced nature of these data.

3.5 Exploring known issues of crowdsourced data within Place Pulse

In section 2, we mentioned a few issues that may be present in crowdsourced data, and which are important to keep in mind when ussing these data in criminological research. Here we explore whether some of these issues are present in data recorded from Place Pulse, and what that might mean for any conclusions we draw from our analyses. We shall keep in mind that the Place Pulse project did not record information about participants' demographic characteristics, and thus we cannot direclty examine the self-selection biases that may affect this dataset (Elliott and Valliant 2017; Chataway et al. 2017), but the sample's self-selection bias should be checked when possible. For instance, the first edition of Place Pulse, which was designed to record some demographic variables from participants, identified that 78.3% of those who reported their sex were males, and only 21.7% were females, and the median self-reported age was 28 years (Salesses, Schechtner, and Hidalgo 2013). Here we examine data from Place Pulse 2.0, which does not record social and demographic variables from participants, but we will examine whether our sample is affected by other issues such as participation inequality, under-representation of certain areas and participation decrease.

3.5.1 Participation inequality ('supercontributors')

Crowdsourced data tend to be affected by a few number of supercontributors that produce most votes (Dubey et al. 2016; Solymosi, Bowers, and Fujiyama 2017). In order to check if our dataset is affected by this, we first need to create a new dataframe showing the number of votes that each study participant had made. To do this, we will use the `group_by()` and `summarise()` functions from the `dplyr` library:

```
voter <- pp_data %>%
  group_by(voter_uniqueid) %>% # create groups based on users unique id
  summarise(num_votes = n()) # print the number of votes by user
```

We can have a look at this new dataframe using the `View()` function. If we do this, we observe that we have some very active participants. The top participant, for instance, has made 7168 votes on places. That is some very prolific participation. On the other hand, we can also see that 7494 of the participants made only one vote. We are definitely seeing signs of participation inequality in these data.

In fact, we can examine how many votes are produced by these 'supercontributors'. For example, we can assess the proportion of votes made by the top 1% of voters. We can do this using the `subset()` and `quantile()` functions:

Least safe place among Atlanta places rated by at least 20 Place Pulse users

Safest place among Atlanta places rated by at least 20 Place Pulse users

Figure 2: Figure 2: Least safe and safest places rated by at least 20 Place Pulse users

```r
# subset top 1% of most prolific participants
top_1percent <- subset(voter, num_votes > quantile(num_votes, prob = 1 - 1/100))
```

We see that this new dataframe contains 954 people, who are our top 1% contributors to the Place Pulse dataset. We will now examine how much of the total number of votes are generated by the top 1% of users:

```r
# proportion of votes by top 1% participants
sum(top_1percent$num_votes) / sum(voter$num_votes) * 100
```

```
## [1] 17.87468
```

That is a lot: 17.87% of the votes are made by the top 1% of contributors. We can also compute the proportion of votes made by the top 10% and 25% of participants. The top 10% contributors are responsible for 46.06% of votes, and the top 25% users contribute the 66.20% of all votes.

3.5.2 Quantifying participation inequality

One way to quantify the extent to which participation inequality exists in our data is by using a Gini index, and visualizing it using a Lorenz curve. The Gini index (or Gini ratio) is a measure of statistical dispersion intended to measure inequality (Gastwirth 1972). Although it is generally used to examine income inequality, it has also been frequently used to assess participation inequality in crowdsourcing platforms (see Solymosi, Bowers, and Fujiyama 2017; Solymosi and Bowers 2018). Similarly, the Lorenz curve is a visual representation of inequality. For this we will need the `ineq` library (Zeileis and Kleiber 2015), and we may have to install it if we do not currently have it in our `R` system:

```r
install.packages("ineq")
```

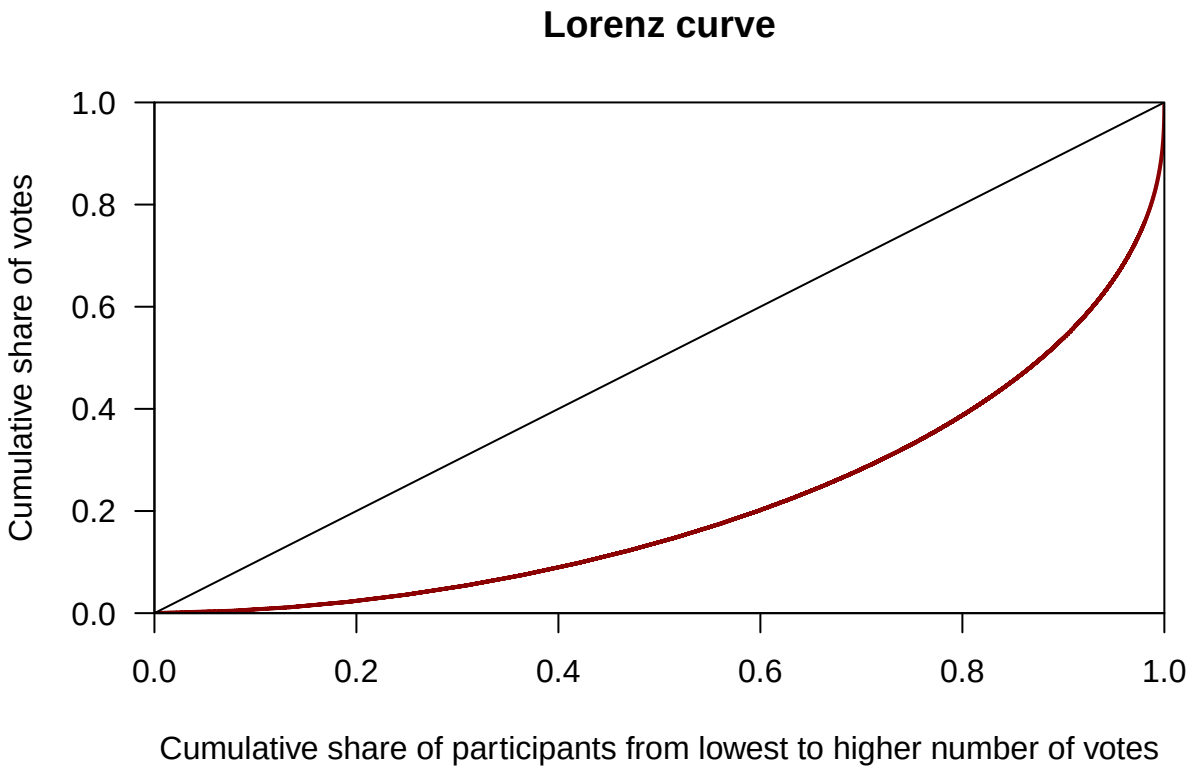
Then we can load this library and calculate the index using the `Gini()` function:

```r
Gini(voter$num_votes) # print Gini index
```

```
## [1] 0.5777568
```

A Gini index score of 0 represents perfect equality (everyone makes equal number of votes), while 1 shows perfect inequality (only one person making every single vote). Our answer of 0.58 shows some serious inequality. To put this into context, in 2017, according to the OECD, income inequality in the United States showed a Gini coefficient of 0.39. To visualize this we can use a Lorenz curve using the `plot()` and `Lc()` functions:

```r
plot(Lc(voter$num_votes), # plot Lorenz curve
     xlab = "Cumulative share of participants from lowest to higher number of votes",
     ylab = "Cumulative share of votes", col = "darkred", lwd = 2)
```



The Lorenz curve (red line) shows how the top few percent of users contribute the majority of the reports. If we had perfect equality, we would expect to see the red line align perfectly with the black line with the slope of 1. With this information we can now quantify how severe the participation inequality is in our data, and compare with other crowdsourced data for context and understanding.

3.5.3 Under-representation of certain areas

It is also important to consider variation in the sample size of number of votes in each census tract. We can analyse if certain areas are under-represented in our dataset by using the `summary()` function.

```
summary(atl_pp_wins$num_votes)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     2.0    87.0   152.0   192.9   250.5  1028.0
```

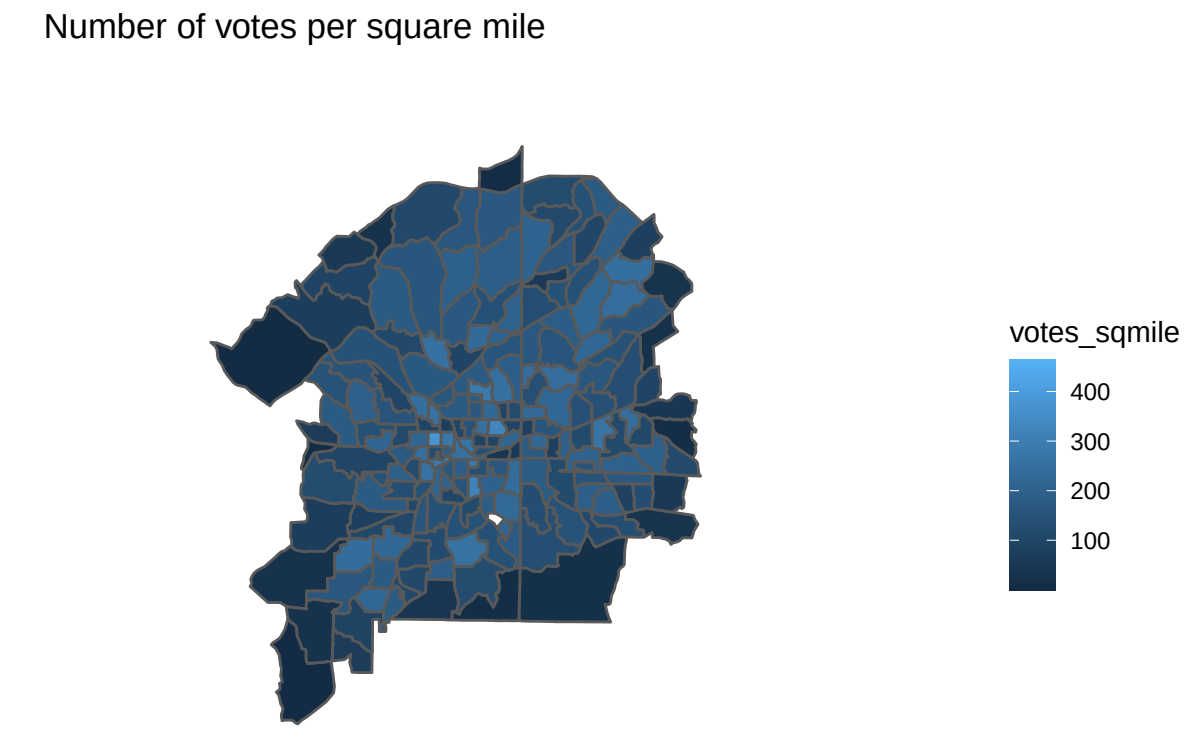
Whereas the average sample size per area is quite large, 192.86, some tracts are clearly over-represented (the maximum number of votes is 1028) and others suffer from small representation (the minimum number of votes is only 2).

We may also want to know which areas suffer from severe under-representation in our dataset. We use the function `mutate()` to compute the number of votes divided by square miles in each census tract.

```
# compute new column of number of votes divded by square miles
atl_pp_wins <- atl_pp_wins %>%
  mutate(votes_sqmile = num_votes / SQ_MILES)
```

Then we can visualize the geographic distribution of the number of votes per census tract using the same code shows above to map perceptions of safety.

```
ggplot(data = atl_pp_wins) +
   ggtitle("Number of votes per square mile") +
   geom_sf(aes(fill = votes_sqmile)) +
   coord_sf(xlim = c(-84.7, -84), ylim = c(33.5, 34), expand = FALSE) +
   theme_void()
```



We see that areas in the city center tend to have larger number of votes and are therefore well represented, whereas tracts in surrounding areas suffer from smaller sample sizes. Estimates of perceived safety in under-represented areas are likely to be affected by a small number of responses and may suffer from low precision. In order to increase the reliability of estimates produced from crowdsourced data for areas with small sample sizes, some researchers suggest using resampling and model-based techniques (see Buil-Gil, Solymosi, and Moretti 2020).

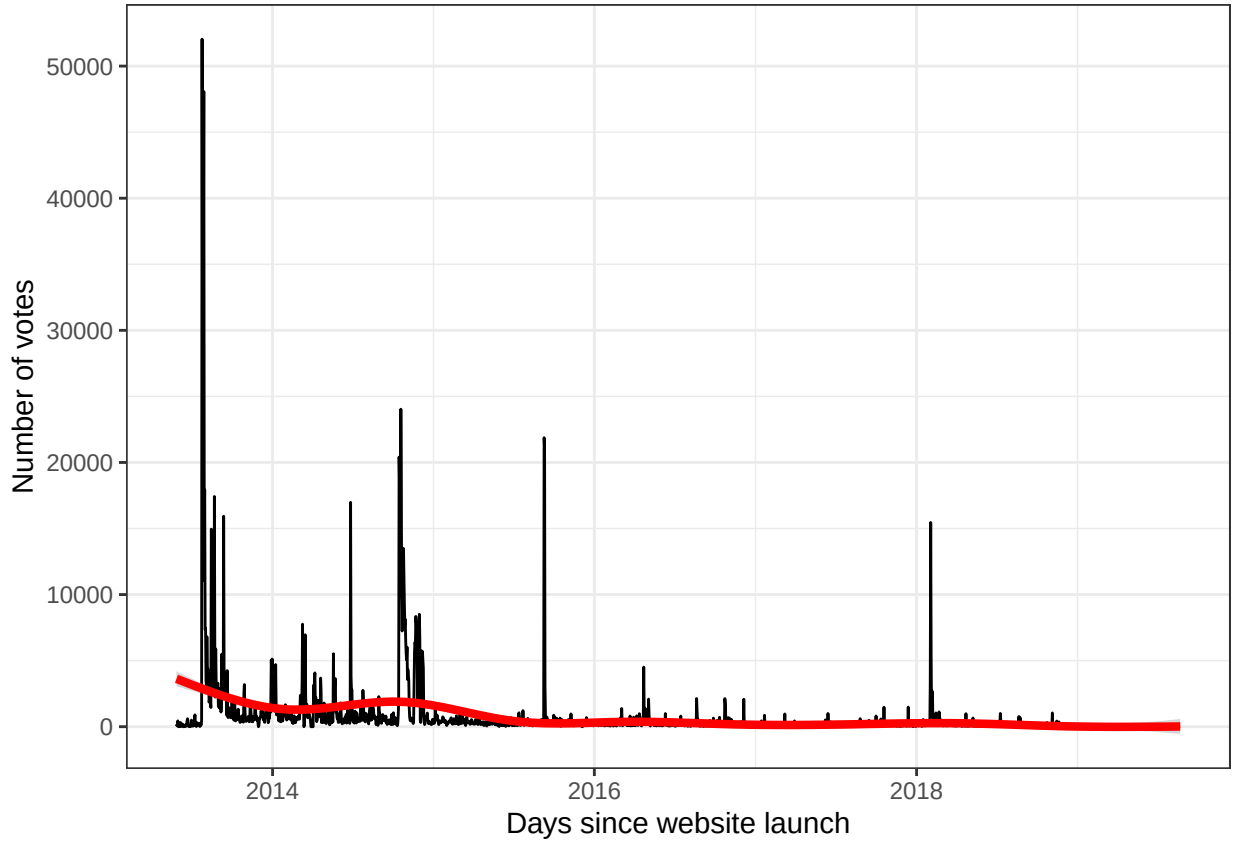
3.5.4 Participation decrease

Finally, some researchers have identified that the number of users of crowdsourcing projects decreases over time: whereas the number of participants tends to be large during the first few days, users lose interest in the project if they do not obtain clear short-term benefits from using it (see Blom et al. 2010; Solymosi et al. 2020). We can also expore whether our Place Pulse dataset is affected by participation decrease.

We will use various functions from `dplyr` and `ggplot2` packages (seen above) to visualize the number of votes since the Place Pulse website launch until its closure. But we also need to use other key functions: *(i)* the `ymd()` function from `lubricate` package is used to transform the dates in which votes took place

into `Date` objects (Spinu, Grolemund, and Wickham 2020), and *(ii)* the `complete()` function from `tidyr` package is used to turn implicit missing dates into explicit missing dates and create a timeline of dates with and without votes (Wickham and Henry 2020).

```
by_day <- pp_data %>%
          mutate(day = ymd(day)) %>% # create column by transforming into 'Date' object
          group_by(day) %>% # create groups by days
          summarise(num_votes = n()) %>% # count number of votes by day
          complete(day = seq.Date(min(day), max(day), by = "day")) %>% # complete all days
          mutate(num_votes = replace_na(num_votes, 0)) # create new column: votes by day

ggplot(by_day, aes(x = day, y = num_votes)) +
       geom_line() +
       geom_smooth(lwd = 1.5, col = "red") +
       theme_bw() +
       xlab("Days since website launch") +
       ylab("Number of votes")
```



The number of votes within the Place Pulse platform clearly decreased over time, but we also observe some peaks even years after the launching of the project. Some of these peaks match the dates of key publications using Place Pulse data and media releases, which shows that participation in crowdsoucing projects can be enhanced by periodic campaigns. For example, we observe a large peak beginning on July 24th 2013, date in which Salesses, Schechtner, and Hidalgo (2013) published their paper and the Massachusetts Institute of Technology published a news article about the Place Pulse platform on their website: http://news.mit.edu/2013/quantifying-urban-perceptions-0724. We also observe another peak of participation beginning on October 15th 2014, just after the publication of Harvey (2014) Master's thesis

about how to automate the study of the characteristics of streetscape skeletons and urban perceptions from Place Pulse data.

Conclusions

The open data movement has provoked a revolution in social research methods, and will continue changing the way in which many social issues are researched, understood and managed. Digital technologies enable large volumes of data to become available for social researchers and data scientists, and crowdsourcing is becoming a key source of data to analyze and map social phenomena such as crime (Bendler et al. 2014) and perceptions of space and safety (Solymosi, Bowers, and Fujiyama 2015; Solymosi and Bowers 2018). In this chapter we have described and explored the main strengths and weaknesses of using crowdsourced data for criminological research. Specifically, we have obtained access to a large dataset of more than 1.5 million votes about urban perceptions recorded from the Place Pulse project (Salesses, Schechtner, and Hidalgo 2013), selected a sample of more than 37,000 votes of perceived safety for Atlanta, and studied the spatial distribution of perceptions of space and safety at a census tract level in this city. We have also shown how these data can be utilized to identify places assessed by participants as very safe or very unsafe; places in which researchers can then conduct observation to study those environmental features that make citizens feel fear of crime (Fisher and Nasar 1992).

Although crowdsourcing offers advantages over traditional survey methods to study perceptions and emotions about crime, data recorded from crowdsourcing is also affected by certain issues that, if uncontrolled, are likely to affect the validity of data and the reliability and generalizability of research outputs. For instance, we have observed how Place Pulse votes are largely produced by a few number of super-contributors (i.e., participation inequality), there is under-representation of certain areas outside the city center, and the number of votes decreases over time (i.e., participation decrease). These issues have also been observed in data produced from many other crowdsourcing and app-based projects (e.g., Chataway et al. 2017; Solymosi, Bowers, and Fujiyama 2015; Solymosi et al. 2020; Traunmueller, Marshall, and Capra 2015). Other researchers have also highlighted that crowdsourced data tends to be affected by self-selection bias, which explains why males tend to participate more than females, and young persons more than adults (Solymosi and Bowers 2018); but the Place Pulse platform did not record demographic variables from participants and we have not directly assessed this issue here.

Due to the fact that crowdsourced datasets - and non-probability samples in general - may be affected by these potential sources of unrepresentativeness and bias, several researchers are exploring new techniques to enable obtaininig reliable research outputs. Elliott and Valliant (2017), for example, present different methods to compute individual pseudo-sampling weights and adjust non-probability samples to target populations; Arbia et al. (2018) have developed a method to delete spatial outliers and calculate weights to adjust non-probability samples to optimal spatial samples; and Buil-Gil, Solymosi, and Moretti (2020) investigate the use of resampling and model-based small area estimation techniques to allow producing reliable estimates at detailed spatial scales from crowdsourced data. Academics and practitioners will benefit from methods to mitigate the sources of bias in crowdsourced data, which may allow obtaining more precise and reliable - but also cheaper - findings and devise new explanations of crime, antisocial behavior and emotions about crime. In the context of crime analysis, bias-corrected crowdsourced data may become a key tool to understand crime patterns, anticipate crime trends and even provide assistance to police investigations (Bendler et al. 2014; Nhan, Huey, and Broll 2017).

Authors bios

David Buil-Gil is a Research Fellow at the Department of Criminology of the University of Manchester, UK, and a member of the Cathie Marsh Institute for Social Research at this same university. His research interests cover small area estimation applications in criminology, environmental criminology, crime mapping, emotions about crime, crime reporting, new methods for data collection and open data.

Reka Solymosi is a Lecturer in Quantiative Methods at the Department of Criminology of the University of Manchetser, UK, with interests in data analysis and visualization, crowdsourcing, rstats, fear of crime, transport, and collecting data about everyday life. As a former crime analyst, she is interested in practical applications to research and solving everyday problems with data.

Acknowledgments

The authors would like to thank César A. Hidalgo for providing access to the data used in this book chapter.

References

Arbia, G., G. Solano-Hermosilla, F. Micale, V. Nardelli, and G. Genovese. 2018. "Post-Sampling Crowdsourced Data to Allow Reliable Statistical Inference: The Case of Food Price Indices in Nigeria." Open Conference Systems.

Bendler, J., T. Brandt, S. Wagner, and D. Neumann. 2014. "Investigating Crime-to-Twitter Relationships in Urban Environments - Facilitating a Virtual Neighborhood Watch." Association for Information Systems.

Birenboim, A. 2016. "New Approaches to the Study of Tourist Experiences in Time and Space." *Tourism Geographies* 18 (1).

Blom, J., D. Viswanathan, J. Go, M. Spasojevic, K. Acharya, and R. Ahonius. 2010. "Fear and the City - Role of Mobile Services in Harnessing Safety and Security in Urban Contexts." Association for Computing Machinery.

Brabham, D. C. 2008. "Crowdsourcing as a Model for Problem Solving: An Introduction and Cases." *Convergence: The International Journal of Research into New Media Technologies* 14 (1).

Buil-Gil, D., A. Moretti, N. Shlomo, and J. Medina. 2019. "Worry About Crime in Europe: A Model-Based Small Area Estimation from the European Social Survey." *European Journal of Criminology*.

Buil-Gil, D., R. Solymosi, and A. Moretti. 2020. "Non-Parametric Bootstrap and Small Area Estimation to Mitigate Bias in Crowdsourced Data: Simulation Study and Application to Perceived Safety." In *Big Data Meets Survey Science*, edited by C. Hill, P. Biemer, T. Buskirk, L. Japec, A. Kirchner, S. Kolenikov, and Lyberg L. John Wiley & Sons Ltd.

Castro-Toledo, F. J., J. O. Perea-García, R. Bautista-Ortuño, and P. Mitkidis. 2017. "Influence of Environmental Variables on Fear of Crime: Comparing Self-Report Data with Physiological Measures in an Experimental Design." *Journal of Experimental Criminology* 13.

Chamberlain, S., and A. Teucher. 2020. *Geojsonio: Convert Data from and to 'Geojson' or 'Topojson'*. https://CRAN.R-project.org/package=geojsonio.

Chataway, M. L., T. C. Hart, R. Coomber, and C. Bond. 2017. "The Geography of Crime Fear: A Pilot Study Exploring Event-Based Perceptions of Risk Using Mobile Technology." *Applied Geography* 86.

Dubey, A., N. Naik, D. Parikh, R. Raskar, and C. A. Hidalgo. 2016. "Deep Learning the City: Quantifying Urban Perception at a Global Scale." In *Computer Vision - Eccv 2016*, edited by B. Leibe, J. Matas, N. Sebe, and M. Welling, 196–212. Cham: Springer.

Elliott, M. R., and R. Valliant. 2017. "Inference for Nonprobability Samples." *Statistical Science* 32 (2).

Erete, S., L. Nicole, J. Mumm, A. Boussayoud, and I. F. Ogbonnaya-Obguru. 2016. ""That Neighborhood Is Sketchy!": Examining Online Conversations About Social Disorder in Transitioning Neighborhoods." Association for Computing Machinery.

Fisher, B. S., and J. L. Nasar. 1992. "Fear of Crime in Relation to Three Exterior Site Features." *Environment and Behavior* 24 (1).

Gabriel, U., and W. Greve. 2003. "The Psychology of Fear of Crime. Conceptual and Methodological Perspectives." *British Journal of Criminology* 43.

Gastwirth, J. L. 1972. "The Estimation of the Lorenz Curve and Gini Index." *The Review of Economics and Statistics* 54 (3).

Goodchild, M. F. 2007. "Citizens as Sensors: The World of Volunteered Geography." *GeoJournal* 69.

Goodchild, M. F., and J. A. Glennon. 2010. "Crowdsourcing Geographic Information for Disaster Response: A Research Frontier." *International Journal of Digital Earth* 3 (3).

Gómez, F., A. Torres, J. Galvis, J. Camargo, and O. Martínez. 2016. "Hotspot Mapping for Perception of Security." IEEE.

Hale, C. 1996. "Fear of Crime: A Review of the Literature." *International Review of Victimology* 4 (2).

Hamilton, M., F. Salim, E. Cheng, and S. L. Choy. 2011. "Transafe: A Crowdsourced Mobile Platform for Crime and Safety Perception Management." IEEE.

Harvey, C. W. 2014. "Measuring Streetscape Design for Livability Using Spatial Data and Methods." Master's thesis, The Faculty of the Graduate College, The University of Vermont.

Hecker, S., W. Wicke, M. Haklay, and A. Bonn. 2019. "How Does Policy Conceptualise Citizen Science? A Qualitative Content Analysis of International Policy Documents." *Citizen Science: Theory and Practice* 4 (1).

Hough, M. 2004. "Worry About Crime: Mental Events or Mental States?" *International Journal of Social Research Methodology* 7 (2).

Howe, J. 2006. "The Rise of Crowdsourcing." *Wired Magazine* 14 (6).

Innes, M. 2015. "'Plac-Ing' Fear of Crime." *Legal and Criminological Psychology* 20 (2).

Jackson, J., and I. Gouseti. 2015. "Psychological Proximity and the Construal of Crime: A Commentary on 'Mapping Fear of Crime as a Context-dependent Everyday Experience That Varies in Space and Time'." *Legal and Criminological Psychology* 20 (2).

McNulty, T. L., and S. R. Holloway. 2000. "Race, Crime, and Public Housing in Atlanta: Testing a Conditional Effect Hypothesis." *Social Forces* 79 (2).

Miró-Llinares, F., A. Moneva, and M. Esteve. 2018. "Hate Is in the Air! But Where? Introducing an Algorithm to Detect Hate Speech in Digital Microenvironments." *Crime Science* 7 (15).

Nhan, J., L. Huey, and R. Broll. 2017. "Digilantism: An Analysis of Crowdsourcing and the Boston Marathon Bombings." *British Journal of Criminology* 57 (2).

Pebesma, E. 2020. *Sf: Simple Features for R.* https://CRAN.R-project.org/package=sf.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.

Salesses, P., K. Schechtner, and C. A. Hidalgo. 2013. "The Collaborative Image of the City: Mapping the Inequality of Urban Perception." *PloS One* 8 (7).

Solymosi, R., and K. Bowers. 2018. "The Role of Innovative Data Collection Methods in Advancing Criminological Understanding." In *The Oxford Handbook of Environmental Criminology*, edited by G. J.N. Bruinsma and S. D. Johnson, 210–37. New York: Oxford University Press.

Solymosi, R., K. Bowers, and T. Fujiyama. 2015. "Mapping Fear of Crime as a Context-dependent Everyday Experience That Varies in Space and Time." *Legal and Criminological Psychology* 20 (2).

Solymosi, R., K. J. Bowers, and T. Fujiyama. 2017. "Crowdsourcing Subjective Perceptions of Neighbourhood Disorder: Interpreting Bias in Open Data." *British Journal of Criminology* 58 (4).

Solymosi, R., D. Buil-Gil, L. Vozmediano, and I. Guedes. 2020. "Towards a Place-Based Measure of Fear of Crime: A Systematic Review of App-Based and Crowdsourcing Approaches." *Environment & Behavior.*

Spinu, V., G. Grolemund, and H. Wickham. 2020. *Lubridate: Make Dealing with Dates a Little Easier*. https://cran.r-project.org/web/packages/lubridate/index.html.

Tester, G., E. Ruel, A. Anderson, D. C. Reitzes, and D. Oakley. 2011. "Sense of Place Among Atlanta Public Housing Residents." *Journal of Urban Health: Bulletin of the New York Academy of Medicine* 88 (3).

Traunmueller, M., P. Marshall, and L. Capra. 2015. "Crowdsourcing Safety Perceptions of People: Opportunities and Limitations." In *SocInfo2015: Social Informatics*, edited by T. Y. Liu, C. Scollon, and W. Zhu, 120–35. Cham: Springer.

Warr, M. 2000. "Fear of Crime in the United States: Avenues for Research and Policy." *Criminal Justice* 4 (4).

Welsh, B. C., and D. P. Farrington. 2004. "Surveillance for Crime Prevention in Public Space: Results and Policy Choices in Britain and America." *Criminology & Public Policy* 3 (3).

Wickham, H., W. Chang, L. Henry, T. L. Pedersen, K. Takahashi, C. Wilke, K. Woo, H. Yutani, and D. Dunnington. 2020. *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. https://CRAN.R-project.org/package=ggplot2.

Wickham, H., R. François, L. Henry, and K. Müller. 2020. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, H., and L. Henry. 2020. *Tidyr: Tidy Messy Data*. https://CRAN.R-project.org/package=tidyr.

Williams, M. L., P. Burnap, A. Javed, H. Liu, and S. Ozalp. 2020. "Hate in the Machine: Anti-Black and Anti-Muslim Social Media Posts as Predictors of Offline Racially and Religiously Aggravated Crime." *British Journal of Criminology* 60 (1).

Williams, M. L., P. Burnap, and L. Sloan. 2017. "Crime Sensing with Big Data: The Affordances and Limitations of Using Open-Source Communications to Estimate Crime Patterns." *British Journal of Criminology* 57.

Zeileis, A., and C. Kleiber. 2015. *Ineq: Measuring Inequality, Concentration, and Poverty*. https://CRAN.R-project.org/package=ineq.