In this post, we will analyze government data from the Flemish region in Belgium on A) official crime statistics and B) self-reported feelings of safety among residents of Flanders. We will focus our analysis on two cities in the province of [Flemish Brabant](https://en.wikipedia.org/wiki/Flemish_Brabant): [Leuven](https://en.wikipedia.org/wiki/Leuven) and [Vilvoorde](https://en.wikipedia.org/wiki/Vilvoorde). A key question of this analysis is: do the residents of the safer city (as measured by official government statistics) feel safer and have more pride in their city (according to government polling data)?

**The Data**

*Crime*

Our first set of data consists of official crime figures from the Flemish government, obtained at [this website](http://www.statistiekvlaanderen.be/statistiek-criminaliteit). I chose to download the data adjusted for population size. Specifically, the data record the number of reported incidents per 1000 residents. The types of crime reported are: property damage, theft, physical violence, and overall incidents. As of October 2018, these figures were available for the years 2000-2016.

*Subjective Feelings About the Cities*

The Flemish government regularly conducts a survey (called the “[Stadsmonitor](https://www.gemeente-en-stadsmonitor.vlaanderen.be/wat-zijn-de-gemeente-en-de-stadsmonitor" \t "_blank)“) to monitor how the residents of Flanders feel about many different topics, including how people feel about the city or town where they live. Summary data are available by question on the [official survey website](https://www.gemeente-en-stadsmonitor.vlaanderen.be/naar-de-cijfers/brondata-en-metadata). In this blog post, we will analyze answers from 2017 on the following questions: insecurity, problems, vandalism and pride in one’s city

*Open Data and Code*

I’m increasingly sharing data and code for blog posts in the hopes that this will be interesting or useful for readers.

**Data Preparation**

*Crime Data*

The code below reads the raw crime data file and first does some basic cleaning. These files were only available in an Excel format, and were not meant to be machine readable, so it takes some effort to manipulate the data to work with them in statistical analysis programs like R.

I then produce an additional set of crime statistics for Leuven. Because Leuven is a university city, there are a number of students who live here during the year, but who are not officially registered as residents of the city. Essentially, these “invisible” residents live in the city, potentially commit or are victims of crime, but are not counted as residents and so do not factor into the calculation of crime per 1000 residents. In my analysis below, I’m assuming that there are an additional 30% “invisible” residents, and therefore divide all of the crime figures by 1.3. I call these data “Leuven student assumptions” because they are adjustments to the actual figures, and not the official statistics themselves.

I then convert the data from its original wide format to a long format, as **ggplot2** requires data in the long format for plotting. The code to read the data, produce the adjustment for students, and transform the data is below:

# load the packages we'll need  
library(openxlsx)  
library(ggplot2)  
library(plyr);

library(dplyr)  
library(tidyr)  
library(likert)  
  
# define path to the data  
in\_dir <- 'C:\\Folder\\'  
  
# function to read in crime data  
read\_crime\_data <- function(file\_name\_f, in\_dir\_f){  
 # read the Excel file  
 crime\_raw\_f <- read.xlsx(paste0(in\_dir\_f, file\_name\_f))  
 # remove weird rows with no data  
 crime\_raw\_f <- crime\_raw\_f[1:12,]  
 # make the column name city (translate from Dutch to English)  
 names(crime\_raw\_f)[names(crime\_raw\_f) == 'Gebied'] <- 'City'  
 # make the Flemish region code also in English  
 crime\_raw\_f$City[crime\_raw\_f$City == "Vlaams Gewest"] <- "Vlaams Gewest (Flemish Region)"  
 return(crime\_raw\_f)  
}  
  
# read in crime data  
crime\_clean <- read\_crime\_data('Leuven Vilvoorde Crime per 1000 inhabitants clean.xlsx', in\_dir)  
  
# make function to adjust for   
# "missing" students  
# living but not registered in Leuven   
divide\_it <- function(x){  
 divided\_figure <- x / 1.3  
 return(divided\_figure)  
}  
  
# create the data for the Leuven student assumptions  
leuven\_w\_students <- crime\_clean %>% filter(City == 'Leuven') %>%  
 # mutate variables in the year columns (crime figures)  
 # we apply our function to each cell (divide by 1.3)  
 mutate\_at(vars(`2000`:`2016`), divide\_it) %>%  
 # change the name of the city to indicate it  
 # reflects our assumptions, not the actual data  
 mutate(City = dplyr::recode(City, 'Leuven' = 'Leuven Student Assumptions' ))  
  
# stack our newly created data on top of original data  
master\_crime <- rbind(crime\_clean, leuven\_w\_students)  
  
# transform from wide to long format  
# ggplot2 needs data in this structure  
crime\_long <- gather(master\_crime, year, value, `2000`:`2016`)

The code produces a dataset (*crime\_long*) that looks like this (only first 10 lines shown):

|  | **City** | **Categorie** | **year** | **value** |
| --- | --- | --- | --- | --- |
| 1 | Leuven | Beschadigen van eigendom (aantal per 1.000 inw.) | 2000 | 14.82 |
| 2 | Leuven | Diefstal en afpersing (aantal per 1.000 inw.) | 2000 | 54.43 |
| 3 | Leuven | Misdr. tegen de lichamelijke integriteit (aantal per 1.000 inw.) | 2000 | 6.17 |
| 4 | Leuven | Criminaliteitsgraad (aantal per 1.000 inw.) | 2000 | 132.81 |
| 5 | Vilvoorde | Beschadigen van eigendom (aantal per 1.000 inw.) | 2000 | 8.47 |
| 6 | Vilvoorde | Diefstal en afpersing (aantal per 1.000 inw.) | 2000 | 44.95 |
| 7 | Vilvoorde | Misdr. tegen de lichamelijke integriteit (aantal per 1.000 inw.) | 2000 | 6.08 |
| 8 | Vilvoorde | Criminaliteitsgraad (aantal per 1.000 inw.) | 2000 | 88.22 |
| 9 | Vlaams Gewest (Flemish Region) | Beschadigen van eigendom (aantal per 1.000 inw.) | 2000 | 9.36 |
| 10 | Vlaams Gewest (Flemish Region) | Diefstal en afpersing (aantal per 1.000 inw.) | 2000 | 37.12 |

Note that we also have data on the entire Flemish region (*Vlaams Gewest*) which is a nice point of comparison – the global average across the territory. The descriptions are all in Dutch, but we’ll use English translations for plotting below.

*Survey Data*

The code below reads in the datasets (there is 1 dataset per question, and each dataset contains summary statistics from many different cities – here we focus on just Leuven and Vilvoorde). Unfortunately, there are slight differences in the formatting of the files between the questions, so I wrote a different function to read in every dataset. The functions read the data, select the observations from Leuven and Vilvoorde in 2017, and make English language column names (the original names are all in Dutch).

###### safety  
read\_safety\_data <- function(file\_name\_f, in\_dir\_f){  
 # read the file, specifying that there are dates  
 safety\_raw\_f <- read.xlsx(paste0(in\_dir, file\_name\_f), detectDates = TRUE)  
 # remove info hanging around at bottom of file  
 safety\_raw\_f <- safety\_raw\_f[!is.na(safety\_raw\_f$Indicator),]  
 # name the columns  
 names(safety\_raw\_f) <- c("Gemeente","Indicator","Item", 'Date', 'Empty', 'Never/Seldom', 'Sometimes', 'Often/Always')  
 # select observations just from 2017 - we only have Vilvoorde for that date  
 safety\_raw\_f <- safety\_raw\_f[safety\_raw\_f$Date == '2017-01-01',]  
 # remove the empty column  
 safety\_raw\_f$Empty <- NULL  
 # replace missings in questionnaire data with zeros  
 safety\_raw\_f[is.na(safety\_raw\_f)] <- 0  
 # select only relevant columns for Leuven and Vilvoorde  
 safety\_raw\_f <- safety\_raw\_f[safety\_raw\_f$Gemeente == 'Leuven' | safety\_raw\_f$Gemeente == 'Vilvoorde',c(1,3,5:7)]  
 # make sure that the "Item" variable represents city vs. neighborhood  
 # (for Likert plot)  
 safety\_raw\_f$Item <- c('Neighborhood', 'City','Neighborhood', 'City')  
 # make city and Item factors  
 # (for Likert plot)  
 safety\_raw\_f$Gemeente <- as.factor(safety\_raw\_f$Gemeente)  
 safety\_raw\_f$Item <- as.factor(safety\_raw\_f$Item)  
 # return the clean dataset  
 return(safety\_raw\_f)  
}  
  
###### problems  
read\_problem\_data <- function(file\_name\_f, in\_dir\_f){  
 # read the file, specifying that there are dates  
 probs\_raw\_f <- read.xlsx(paste0(in\_dir, file\_name\_f), detectDates = TRUE)  
 # remove info hanging around at bottom of file  
 probs\_raw\_f <- probs\_raw\_f[!is.na(probs\_raw\_f$Indicator),]  
 # name the columns  
 names(probs\_raw\_f) <- c("Gemeente","Indicator","Item", 'Date', 'Empty', 'Never/Seldom', 'Sometimes', 'Often/Always')  
 # select observations just from 2017 - we only have Vilvoorde for that date  
 probs\_raw\_f <- probs\_raw\_f[probs\_raw\_f$Date == '2017-01-01',]  
 # remove the empty column  
 probs\_raw\_f$Empty <- NULL  
 # replace missings in questionnaire data with zeros  
 probs\_raw\_f[is.na(probs\_raw\_f)] <- 0  
 # select only relevant columns for Leuven and Vilvoorde  
 probs\_raw\_f <- probs\_raw\_f[probs\_raw\_f$Gemeente == 'Leuven' | probs\_raw\_f$Gemeente == 'Vilvoorde',c(1,3,5:7)]  
 # delete the original item column (this is text on y axis)  
 probs\_raw\_f$Item <- NULL  
 # make city (Gemeente) the "Item"  
 # (for Likert plot)  
 names(probs\_raw\_f)[1] <- 'Item'  
 # return the clean dataset  
 return(probs\_raw\_f)  
}  
  
####### vandalism  
read\_vandalism\_data <- function(file\_name\_f, in\_dir\_f){  
 # read the file, specifying that there are dates  
 vand\_raw\_f <- read.xlsx(paste0(in\_dir, file\_name\_f), detectDates = TRUE)  
 # remove info hanging around at bottom of file  
 vand\_raw\_f <- vand\_raw\_f[!is.na(vand\_raw\_f$Indicator),]  
 # name the columns  
 names(vand\_raw\_f) <- c("Gemeente","Indicator","Item", 'Date', 'Empty', 'Never/Seldom', 'Sometimes', 'Often/Always')  
 # select observations just from 2017 - we only have Vilvoorde for that date  
 vand\_raw\_f <- vand\_raw\_f[vand\_raw\_f$Date == '2017' & vand\_raw\_f$Item == 'Vernieling straatmeubilair',]  
 # remove the empty column  
 vand\_raw\_f$Empty <- NULL  
 # replace missings in questionnaire data with zeros  
 vand\_raw\_f[is.na(vand\_raw\_f)] <- 0  
 # select only relevant columns for Leuven and Vilvoorde  
 vand\_raw\_f <- vand\_raw\_f[vand\_raw\_f$Gemeente == 'Leuven' | vand\_raw\_f$Gemeente == 'Vilvoorde',c(1,3,5:7)]  
 # delete the original item column (this is text on y axis)  
 vand\_raw\_f$Item <- NULL  
 # make city (Gemeente) the "Item"  
 # (for Likert plot)  
 names(vand\_raw\_f)[1] <- 'Item'  
 # return the clean dataset  
 return(vand\_raw\_f)  
}  
  
######### pride in one's city  
read\_pride\_data <- function(file\_name\_f, in\_dir\_f){  
 # read the file, specifying that there are dates  
 pride\_raw\_f <- read.xlsx(paste0(in\_dir, file\_name\_f), detectDates = TRUE)  
 # remove info hanging around at bottom of file  
 pride\_raw\_f <- pride\_raw\_f[!is.na(pride\_raw\_f$Indicator),]  
 # name the columns  
 names(pride\_raw\_f) <- c("Gemeente","Indicator","Item", 'Date', 'Empty', 'Mostly/Completely disagree',   
 'Neither agree nor disagree', 'Mostly/Completely agree')  
 # select observations just from 2017 - we only have Vilvoorde for that date  
 pride\_raw\_f <- pride\_raw\_f[pride\_raw\_f$Date == '2017-01-01',]  
 # remove the empty column  
 pride\_raw\_f$Empty <- NULL  
 # replace missings in questionnaire data with zeros  
 pride\_raw\_f[is.na(pride\_raw\_f)] <- 0  
 # select only relevant columns for Leuven and Vilvoorde  
 pride\_raw\_f <- pride\_raw\_f[pride\_raw\_f$Gemeente == 'Leuven' | pride\_raw\_f$Gemeente == 'Vilvoorde',c(1,3,5:7)]  
 # delete the original item column (this is text on y axis)  
 pride\_raw\_f$Item <- NULL  
 # make city (Gemeente) the "Item"  
 # (for Likert plot)  
 names(pride\_raw\_f)[1] <- 'Item'  
 # return the clean dataset  
 return(pride\_raw\_f)  
}  
  
# read in the data  
safety\_clean <- read\_safety\_data('SAMEN\_onveiligheid.xlsx', in\_dir)  
probs\_clean <- read\_problem\_data('SAMEN\_problemen.xlsx', in\_dir)  
vandal\_clean <- read\_vandalism\_data('SAMEN\_vandalisme.xlsx', in\_dir)  
pride\_clean <- read\_pride\_data('WON\_fierheid.xlsx', in\_dir)

Our functions are all executed in a single line, and return a cleaned dataset for each of the four questions we will examine.

For illustrative purposes, the safety dataset is shown below. The answers are in response to the following question: *How often do you feel unsafe in your neighborhood and city?*

|  | **Gemeente** | **Item** | **Never/Seldom** | **Sometimes** | **Often/Always** |
| --- | --- | --- | --- | --- | --- |
| 1 | Leuven | Neighborhood | 83.79 | 13.16 | 3.05 |
| 2 | Leuven | City | 74.76 | 20.88 | 4.36 |
| 3 | Vilvoorde | Neighborhood | 66.04 | 24.05 | 9.91 |
| 4 | Vilvoorde | City | 46.67 | 31.22 | 22.11 |

**Visualization**

*Crime Data*

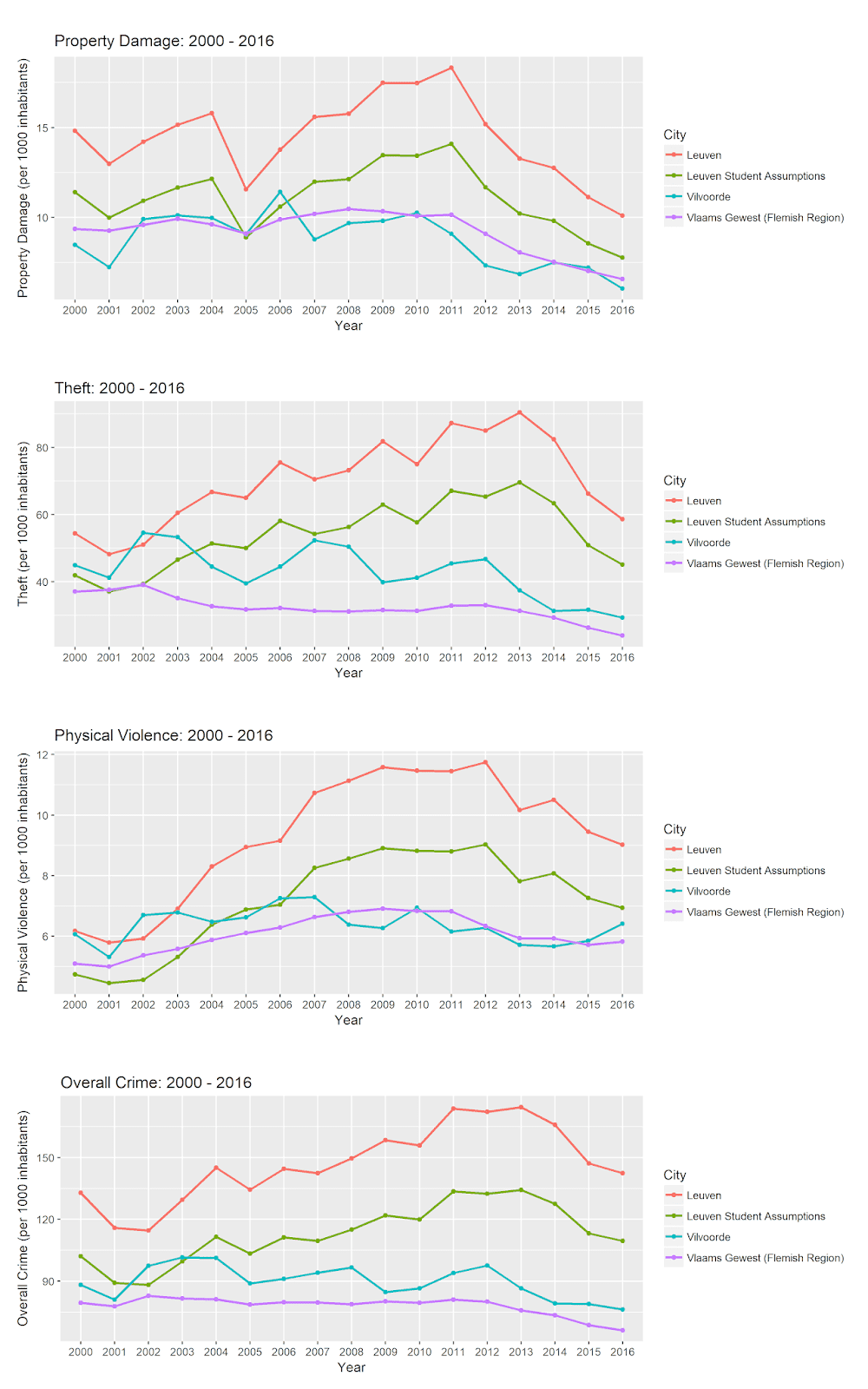
We will produce one graph for each outcome, and assemble the graphs into a single plot. For each outcome, we will plot four different lines, displaying the crime figures across the years for Leuven (“Leuven” in the plots below), Leuven with our assumptions about the student population (assuming 30% “invisible” student residents as described above, “Leuven Student Assumptions” in the plots below), Vilvoorde, and the Vlaams Gewest (average for the entire Flemish region).

The code below stores each single graph in an object, and uses the **gridExtra** package to plot them all in a single plot.

The code looks like this:

# property damage  
prop\_damage\_plot <- crime\_long %>% filter(Categorie == 'Beschadigen van eigendom (aantal per 1.000 inw.)') %>%   
 ggplot(aes(x = year, y = value, color = City, group = City)) +   
 geom\_line(aes(color = City), size = 1) +   
 geom\_point() +   
 labs(x = 'Year', y = 'Property Damage (per 1000 inhabitants)',   
 title = 'Property Damage: 2000 - 2016')   
  
# theft  
theft\_plot <- crime\_long %>% filter(Categorie == 'Diefstal en afpersing (aantal per 1.000 inw.)') %>%   
 ggplot(aes(x = year, y = value, color = City, group = City)) +   
 geom\_line(aes(color = City), size = 1) +   
 geom\_point() +   
 labs(x = 'Year', y = 'Theft (per 1000 inhabitants)',   
 title = 'Theft: 2000 - 2016')   
  
# physical violence  
phys\_violence\_plot <- crime\_long %>% filter(Categorie == 'Misdr. tegen de lichamelijke integriteit (aantal per 1.000 inw.)') %>%   
 ggplot(aes(x = year, y = value, color = City, group = City)) +   
 geom\_line(aes(color = City), size = 1) +   
 geom\_point() +   
 labs(x = 'Year', y = 'Physical Violence (per 1000 inhabitants)',   
 title = 'Physical Violence: 2000 - 2016')   
  
# overall crime  
overall\_crime\_plot <- crime\_long %>% filter(Categorie == 'Criminaliteitsgraad (aantal per 1.000 inw.)') %>%   
 ggplot(aes(x = year, y = value, color = City, group = City)) +   
 geom\_line(aes(color = City), size = 1) +   
 geom\_point() +   
 labs(x = 'Year', y = 'Overall Crime (per 1000 inhabitants)',   
 title = 'Overall Crime: 2000 - 2016')   
  
  
# plot multiple graphs with grid.arrange  
# from gridExtra package  
# we give a bit of extra space at the bottom  
# of each graph  
library(gridExtra)  
margin = theme(plot.margin = unit(c(1,.5,.5,.5), "cm"))  
grid.arrange(grobs = lapply(list(prop\_damage\_plot, theft\_plot, phys\_violence\_plot, overall\_crime\_plot), "+", margin), nrow = 4)

Which produces the following plot.

[](https://i2.wp.com/4.bp.blogspot.com/-9xiaC1lMeKs/W8Nd2errDWI/AAAAAAAAAcg/Z661IUh0LiQVxd6UYxd360-hyPpLLAmYQCLcBGAs/s1600/crime_plot.png?ssl=1)

The official crime statistics for Leuven are higher than the other points of comparison for every outcome after 2003. Even with our assumptions about the “invisible” residents, Leuven consistently has higher crime rates than Vilvoorde and the Flemish region average. Vilvoorde sticks close to the Flemish average for property damage and physical violence, but is higher than average for theft and overall crime rates.

*Questionnaire / Self-Report Data*

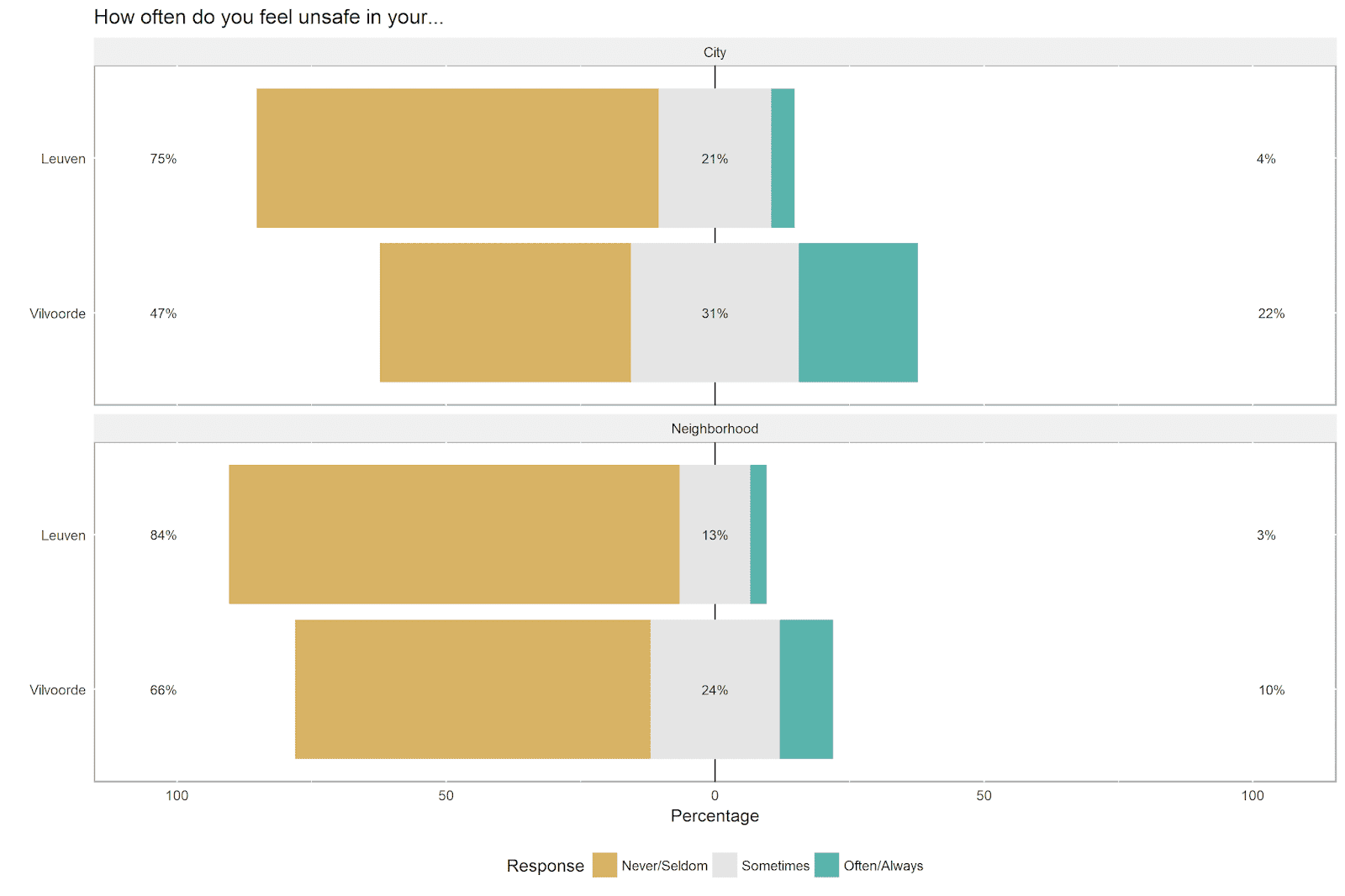
Subjective Feelings of Safety

We will first take a look at subjective feelings of safety in one’s city and neighborhood. We will use the **likert** package to plot the responses to the questionnaires. The Likert package is great for producing stacked bar charts, a common visualization of responses to questionnaire data. The format of the data from the Stadsmonitor is also suited to this type of visual representation.

We first create a Likert scale object from the safety dataset we created above. Because the same question was asked in regard to both one’s neighborhood and one’s city, we will use a grouping in our plot. This will allow us to display the responses to both areas with a correct sub-heading for both neighborhood and city.

The code below creates the Likert object and produces the plot.

# make likert object  
# for safety we use a grouping: neighborhood vs. city  
# (questions were asked about both)  
safety\_likert <- likert(summary = safety\_clean, grouping = safety\_clean[,1])   
# make the plot  
plot(safety\_likert, group.order = c('Leuven', 'Vilvoorde')) + ggtitle('How often do you feel unsafe in your...')

[](https://i2.wp.com/4.bp.blogspot.com/-xAbcvEz63qg/W8NkxAAReZI/AAAAAAAAAcs/PBFBQBTNzIQk2PWhBz1qdPgoqNZmsFWdACLcBGAs/s1600/subjective_safety.png?ssl=1)

Leuven respondents rarely report feeling unsafe in their city (only 4% say they feel unsafe *always* or *often*) and neighborhood (only 3%). The figures are considerably higher for Vilvoorde. Fully 22% of respondents in Vilvoorde say that they *always* or *often* feel unsafe in their city, and 10% say that they *always* or *often* feel unsafe in their neighborhood.

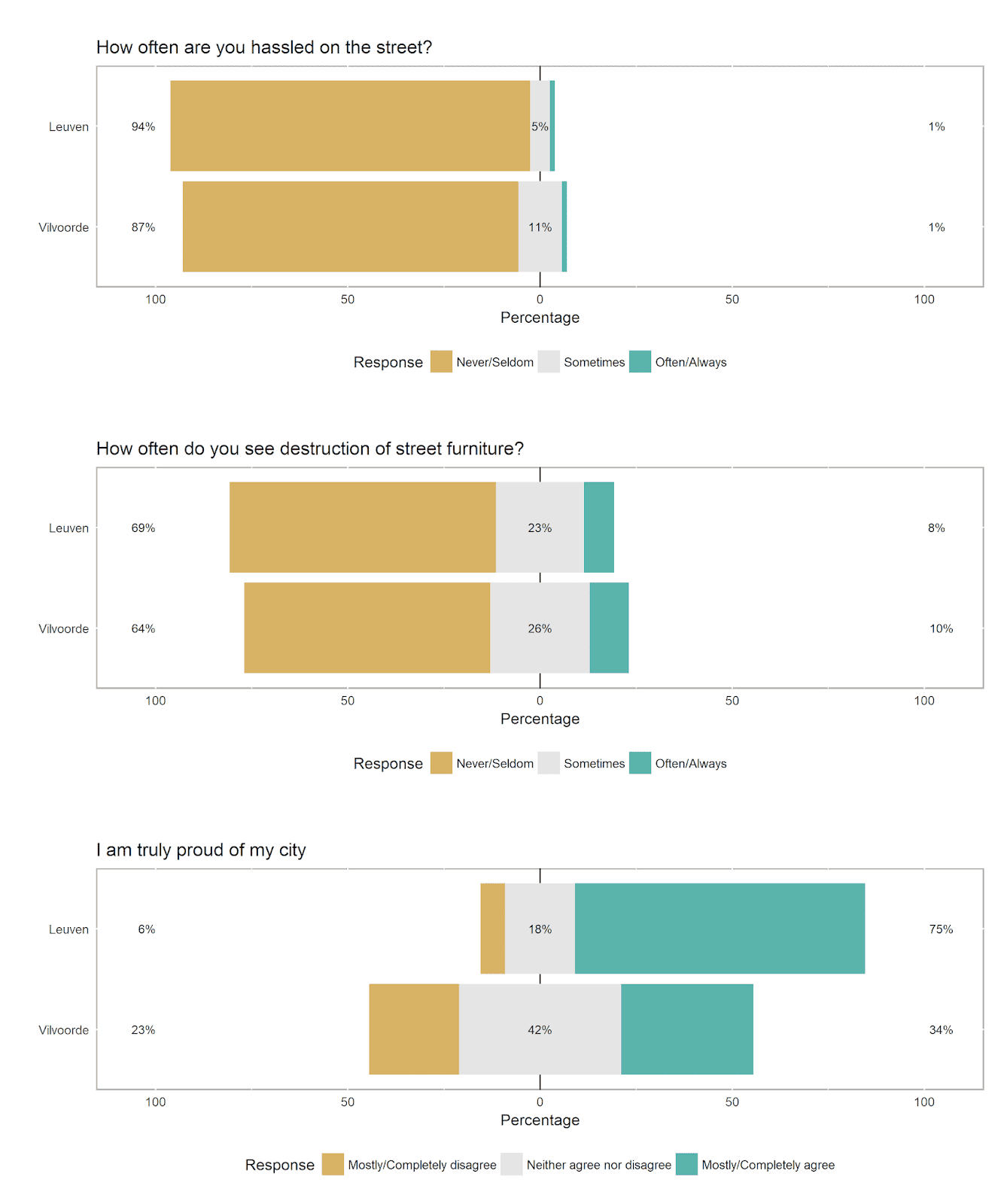
Social Problems and Pride in One’s City

We will next examine questions about two social problems: being hassled on the street and seeing vandalism.We will also look at reports of feelings of pride in one’s city (an overall perception which is no doubt informed by feelings of safety).

We will display these three questions in a single plot, as we did above for the crime data. We first create a Likert graph object for each question (as the questions don’t have the same question structure and response options, it’s better to plot each one by itself). We then use the **gridExtra** package to display all of the graphs in a single plot.

The following code creates the Likert graph objects and displays them in one large plot.

# make likert objects for the other questions  
likert\_problems <- likert(summary = probs\_clean)  
likert\_vand <- likert(summary = vandal\_clean)  
likert\_pride <- likert(summary = pride\_clean)  
  
# first make object for each graph  
prob\_plot <- plot(likert\_problems, group.order = c('Leuven', 'Vilvoorde')) + ggtitle('How often are you hassled on the street?')  
vand\_plot <- plot(likert\_vand, group.order = c('Leuven', 'Vilvoorde')) + ggtitle('How often do you see destruction of street furniture?')  
pride\_plot <- plot(likert\_pride, group.order = c('Leuven', 'Vilvoorde')) + ggtitle('I am truly proud of my city')  
  
# plot multiple graphs with grid.arrange  
# from gridExtra package  
# we give a bit of extra space at the bottom  
# of each graph  
margin = theme(plot.margin = unit(c(1,.5,.5,.5), "cm"))  
grid.arrange(grobs = lapply(list(prob\_plot, vand\_plot, pride\_plot), "+", margin))

[](https://i0.wp.com/4.bp.blogspot.com/-HL-toEiKAFo/W9WvFbhIZ3I/AAAAAAAAAdQ/Xxc8LD0CZ_oPrjUmDd5wjnmvkn6oX8TDgCLcBGAs/s1600/multi_likert_plot.png?ssl=1)

When considering the social problems questions (being hassled on the street and seeing vandalism), the percentage of respondents in Leuven and Vilvoorde who report that they *often* or *always* experience these things is more-or-less the same. The cities distinguish themselves in the middle and lower response categories. More residents in Vilvoorde (vs. Leuven) say that they *sometimes* are hassled on the street and witness vandalism, and fewer residents in Vilvoorde (vs. Leuven) say that they *never* or *seldom* experience these things. In sum, residents of Leuven report fewer social and safety-related problems than residents of Vilvoorde.

The results of the pride question are quite striking. In Leuven, 75% of the respondents are proud of their city versus only 34% in Vilvoorde (less than half the Leuven percentage). Only 6% of Leuven residents are not proud of their city, while this figure is 23% in Vilvoorde.

The self-report data are clear: residents of Leuven feel safer, report fewer social problems and are much more proud of their city when compared with residents of Vilvoorde.

**Summary and Conclusion**

In this post, we examined official crime statistics and survey data on subjective feelings of safety in two Flemish cities: Leuven and Vilvoorde. Interestingly, the official crime statistics and survey data seem to tell two different stories.

On the one hand, Leuven (compared to Vilvoorde) has considerably higher rates of property damage, theft and physical violence, even when assuming that the official statistics are slightly biased because of “invisible” students who are not legally registered as city residents.

On the other hand, people in Leuven say that they feel much safer in their neighborhoods and in their city. Residents of Leuven also report seeing fewer social problems and are prouder of their city.

It’s interesting to see this disconnect between official crime statistics and self-reported feelings of safety. I will admit that I’m not entirely sure what to make of this. Potential explanations for this disconnect include:

1. There are more “invisible” student residents than I have assumed. If this is the case, it is possible that the true population-adjusted crime figures are lower in Leuven than in Vilvoorde (although to close the gap between the cities, there would have to be a very large number of invisible residents in Leuven).
2. Residents in Vilvoorde under-report crime that happens in their city. If this is the case, the actual crime figures in Vilvoorde are much higher, perhaps equal to or greater than those in Leuven (although the similarity between the Vilvoorde figures and those of the Flemish region suggests this is unlikely).
3. The “types” of people who respond to the survey questions are different in Leuven vs. Vilvoorde. If this is the case, the difference in observed feelings of safety and seeing social problems reflects sampling bias, not the true aggregate feelings of city residents. My sense of these data is that, in some contexts, *actual* safety and *feelings* of safety are different things. The overall crime figures in Leuven are not very high, even though they are much higher than the average for the Flemish region (which is, after all, a relatively well-off part of the world). Given the relatively low crime rate overall, perhaps other factors influence how residents feel in the place where they live.

Leuven is a wealthy city in a [wealthy province](https://www.hln.be/nieuws/binnenland/inwoners-uit-vlaams-en-waals-brabant-hebben-de-hoogste-inkomens~a253bf27/). My sense is that people have a common identity as Leuven residents (*Leuvenaars* in Dutch) and are in many ways a homogeneous group (despite a sizeable number of foreign students and academics). Furthermore, the city is doing well economically and has a vision for itself (the university, science and technology spinoffs, along with a large bank and brewery company are all valuable sources of revenue for the city and its residents and seem likely to ensure a solid future in the modern economy).

Vilvoorde, by comparison, is much less wealthy. It is [socially very diverse](https://www.standaard.be/gemeenteprofiel/Vilvoorde), with its residents coming from many different parts of the world. As such, it doesn’t have a single strong source of identity like Leuven, and its diverse residents [do not always mix or know each other very well](http://www.standaard.be/cnt/dmf20180629_03589752).

Given the different economic and social situations of these two cities, it is perhaps understandable that Leuven residents *feel* safer than Vilvoorde residents, even when the actual crime statistics suggest that the opposite is true.