Solution

Martin Arnold   Jens Klenke

11 April, 2022

Optimize code chunks

- When a code chunk is time-consuming to run, you may consider caching it via the chunk option `cache = TRUE`. When the `cache` is turned on, knitr will skip the execution of this code chunk if it has been executed before and nothing in the code chunk has changed since then.

Hide your code chunks by setting a global option

```{r setup, include = FALSE}
knitr::opts_chunk$set(echo = FALSE)
```

Obtain the Crime Incidents Data

First, we download the data and import it to R.

```r
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.8
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(data.table)
```

```
##
## Attaching package: 'data.table'
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
## The following object is masked from 'package:purrr':
##
##     transpose
```

```r
crime <- readr::read_csv(here::here('SoSe_2022/Rmarkdown/Exercise/crime_boston.csv'))
```

```
## Rows: 16909 Columns: 17
```

```
## -- Column specification ----------------------------------------------------
## Delimiter: ","
## chr  (5): OFFENSE_DESCRIPTION, DISTRICT, DAY_OF_WEEK, STREET, Location
## dbl  (9): INCIDENT_NUMBER, OFFENSE_CODE, REPORTING_AREA, SHOOTING, YEAR, MON...
## lgl  (2): OFFENSE_CODE_GROUP, UCR_PART
## dttm (1): OCCURRED_ON_DATE
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
#var_exp <- readxl::read_xlsx(here::here("Rmarkdown","Exercises","rmscrimeincidentfieldexplanation.xlsx
crime_description <- readxl::read_xlsx(here::here('SoSe_2022/Rmarkdown/Exercise/rmscrimeincidentfieldexp
```

Let's have a look at the variable description:

```r
knitr::kable(crime_description)
```

Field Name, Data Type, Required	Description
[incident_num] varchar NOT NULL,	Internal BPD report number
[offense_code][varchar](25) NULL,	Numerical code of offense description
[Offense_Code_Group_Description][varchar](80) NULL,	Internal categorization of [offense_description]
[Offense_Description][varchar](80) NULL,	Primary descriptor of incident
[district] varchar NULL,	What district the crime was reported in
[reporting_area] varchar NULL,	RA number associated with the where the crime was reported from.
[shooting][char] (1) NULL,	Indicated a shooting took place.
[occurred_on] datetime2 NULL,	Earliest date and time the incident could have taken place
[UCR_Part] varchar NULL,	Universal Crime Reporting Part number (1,2, 3)
[street] varchar NULL,	Street name the incident took place

Exploratory data analysis (EDA)

First we want to get an overview of the data.

```r
glimpse(crime)
```

```
## Rows: 16,909
## Columns: 17
## $ INCIDENT_NUMBER     <dbl> 225520077, 222648862, 222201764, 222201559, 222111~
## $ OFFENSE_CODE        <dbl> 3126, 3831, 724, 301, 619, 3126, 801, 611, 619, 32~
## $ OFFENSE_CODE_GROUP  <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
## $ OFFENSE_DESCRIPTION <chr> "WARRANT ARREST - OUTSIDE OF BOSTON WARRANT", "M/V~
## $ DISTRICT            <chr> "D14", "B2", "C6", "D4", "D14", "D4", "C6", "A1", ~
## $ REPORTING_AREA      <dbl> 786, 288, 200, NA, 778, NA, 235, 77, 186, 574, 280~
## $ SHOOTING            <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ OCCURRED_ON_DATE    <dttm> 2022-02-02 00:00:00, 2022-02-05 18:25:00, 2022-01~
## $ YEAR                <dbl> 2022, 2022, 2022, 2022, 2022, 2022, 2022, 2022, 20~
## $ MONTH               <dbl> 2, 2, 1, 3, 2, 3, 2, 2, 1, 1, 1, 2, 1, 2, 4, 4, 4,~
## $ DAY_OF_WEEK         <chr> "Wednesday", "Saturday", "Sunday", "Saturday", "Mo~
## $ HOUR                <dbl> 0, 18, 0, 13, 12, 10, 22, 10, 15, 13, 12, 0, 1, 4,~
## $ UCR_PART            <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
## $ STREET              <chr> "WASHINGTON ST", "WASHINGTON ST", "W BROADWAY", "A~
## $ Lat                 <dbl> 42.34308, 42.32975, 42.34129, 42.33318, 42.34906, ~
```
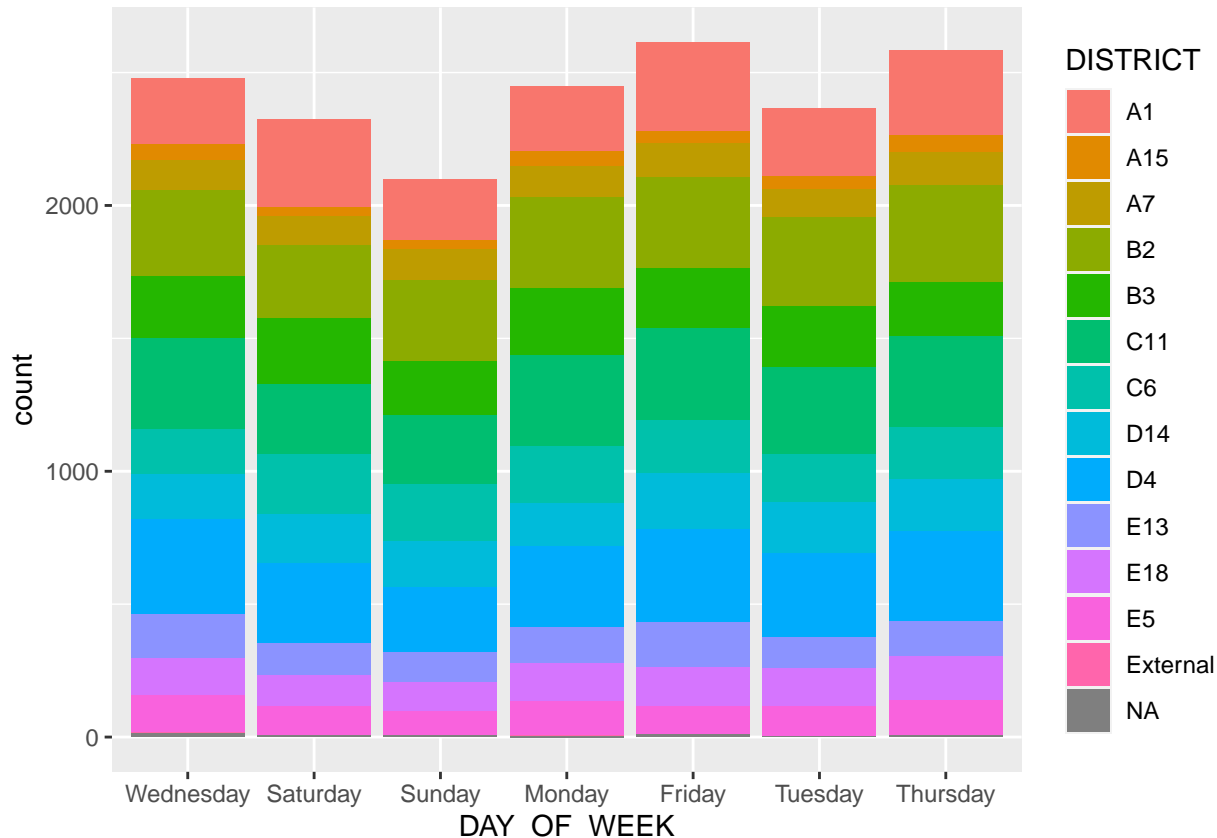
```
## $ Long          <dbl> -71.14172, -71.08454, -71.05468, -71.07394, -71.15~
## $ Location       <chr> "(42.34308127134165, -71.14172267328729)", "(42.32~
```

As we can see, the dataset consists of 16909 Observations in 17 Variables.

Lets focus on the DAY_OC_WEEK.

```
# OFFENSE_CODE should be a factor
crime <- crime %>% mutate(DAY_OF_WEEK = as_factor(DAY_OF_WEEK))

ggplot(crime) +
  geom_bar(aes(x = DAY_OF_WEEK, fill = DISTRICT))
```


Some districts have more crimes than others and the crime is a little bit lazy on sundays.

Nested ordered list for LaTeX output

https://tex.stackexchange.com/questions/78842/nested-enumeration-numbering

1. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum.

    1.1. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum.

2. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum.