

Data Analysis

Donata Buozyte

19/12/2021

Transport Data Tables

Summary and head of all data tables + renaming columns:

```
for (elem in transport_dts){
  print(transport_dt_names[elem])
  names(all_tables[[elem]]) <- gsub(" ", "_", tolower(names(all_tables[[elem]])))
  print(head(all_tables[[elem]]))
  cat("\n")
  print(summary(all_tables[[elem]]))
  cat("\n\n")
}
```

```
## [1] "accidents"
##           id district_name      neighborhood_name
## 1: 2017S008429           Unknown                Unknown
## 2: 2017S007316           Unknown                Unknown
## 3: 2017S010210           Unknown                Unknown
## 4: 2017S006364           Unknown                Unknown
## 5: 2017S004615      Sant Martí el Camp de l'Arpa del Clot
## 6: 2017S007775      Sant Martí el Camp de l'Arpa del Clot
##           street      weekday      month
## 1: Número 27           Friday    October
## 2: Número 3 Zona Franca / Número 50 Zona Franca      Friday September
## 3: Litoral (Besòs)           Friday    December
## 4: Número 3 Zona Franca      Friday      July
## 5: Las Navas de Tolosa      Thursday      May
## 6: Indústria / Trinxant      Wednesday September
##   day hour part_of_the_day mild_injuries serious_injuries victims
## 1:  13    8      Morning           2             0             2
## 2:   1   13      Morning           2             0             2
## 3:   8   21    Afternoon           5             0             5
## 4:  21    2       Night            1             0             1
## 5:  25   14    Afternoon           1             0             1
## 6:  20   12      Morning           1             0             1
##   vehicles_involved longitude latitude
## 1:                 2  2.125624 41.34004
## 2:                 2  2.120452 41.33943
## 3:                 2  2.167356 41.36089
## 4:                 2  2.124529 41.33767
## 5:                 3  2.185272 41.41636
## 6:                 2  2.183245 41.41634
##
##           id      district_name      neighborhood_name
## Length:10339      Length:10339      Length:10339
## Class :character      Class :character      Class :character
## Mode  :character      Mode  :character      Mode  :character
##
```

```

##
##
##      street      weekday      month      day
## Length:10339    Length:10339    Length:10339    Min.   : 1.00
## Class :character Class :character Class :character 1st Qu.: 8.00
## Mode  :character Mode  :character Mode  :character Median :16.00
##                                           Mean  :15.78
##                                           3rd Qu.:23.00
##                                           Max.   :31.00
##      hour      part_of_the_day  mild_injuries  serious_injuries
## Min.   : 0.00    Length:10339    Min.   : 0.000    Min.   :0.00000
## 1st Qu.:10.00    Class :character 1st Qu.: 1.000    1st Qu.:0.00000
## Median :14.00    Mode  :character Median : 1.000    Median :0.00000
## Mean   :13.81                    Mean  : 1.154    Mean  :0.02331
## 3rd Qu.:18.00                    3rd Qu.: 1.000    3rd Qu.:0.00000
## Max.   :23.00                    Max.   :10.000    Max.   :4.00000
##      victims      vehicles_involved  longitude  latitude
## Min.   : 0.000    Min.   : 0.000    Min.   :2.092    Min.   :41.32
## 1st Qu.: 1.000    1st Qu.: 2.000    1st Qu.:2.145    1st Qu.:41.39
## Median : 1.000    Median : 2.000    Median :2.163    Median :41.40
## Mean   : 1.179    Mean   : 1.921    Mean   :2.163    Mean   :41.40
## 3rd Qu.: 1.000    3rd Qu.: 2.000    3rd Qu.:2.180    3rd Qu.:41.41
## Max.   :10.000    Max.   :14.000    Max.   :2.223    Max.   :41.47
##
##
## [1] "bus_stops"
##      code      transport longitude latitude  bus.stop      district.name
## 1: K014 Day bus stop 2.171619 41.41374 BUS -192--    Horta-Guinardó
## 2: K014 Day bus stop 2.134902 41.42022 BUS -124--    Gràcia
## 3: K014 Day bus stop 2.162913 41.42319 BUS -117--    Horta-Guinardó
## 4: K014 Day bus stop 2.163667 41.42290 BUS -117--    Horta-Guinardó
## 5: K014 Day bus stop 2.120212 41.39721 BUS -130--    Sarrià-Sant Gervasi
## 6: K014 Day bus stop 2.138941 41.41108 BUS -131--    Sarrià-Sant Gervasi
##      neighborhood.name
## 1:      el Guinardó
## 2: Vallcarca i els Penitents
## 3:      la Font d'en Fargues
## 4:      la Font d'en Fargues
## 5:      Sarrià
## 6:      el Putxet i el Farró
##
##      code      transport      longitude      latitude
## Length:3162    Length:3162    Min.   :2.056    Min.   :41.32
## Class :character Class :character 1st Qu.:2.141    1st Qu.:41.39
## Mode  :character Mode  :character Median :2.161    Median :41.41
##                                           Mean  :2.160    Mean  :41.41
##                                           3rd Qu.:2.180    3rd Qu.:41.42
##                                           Max.   :2.222    Max.   :41.47
##      bus.stop      district.name  neighborhood.name
## Length:3162    Length:3162    Length:3162
## Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character
##
##

```

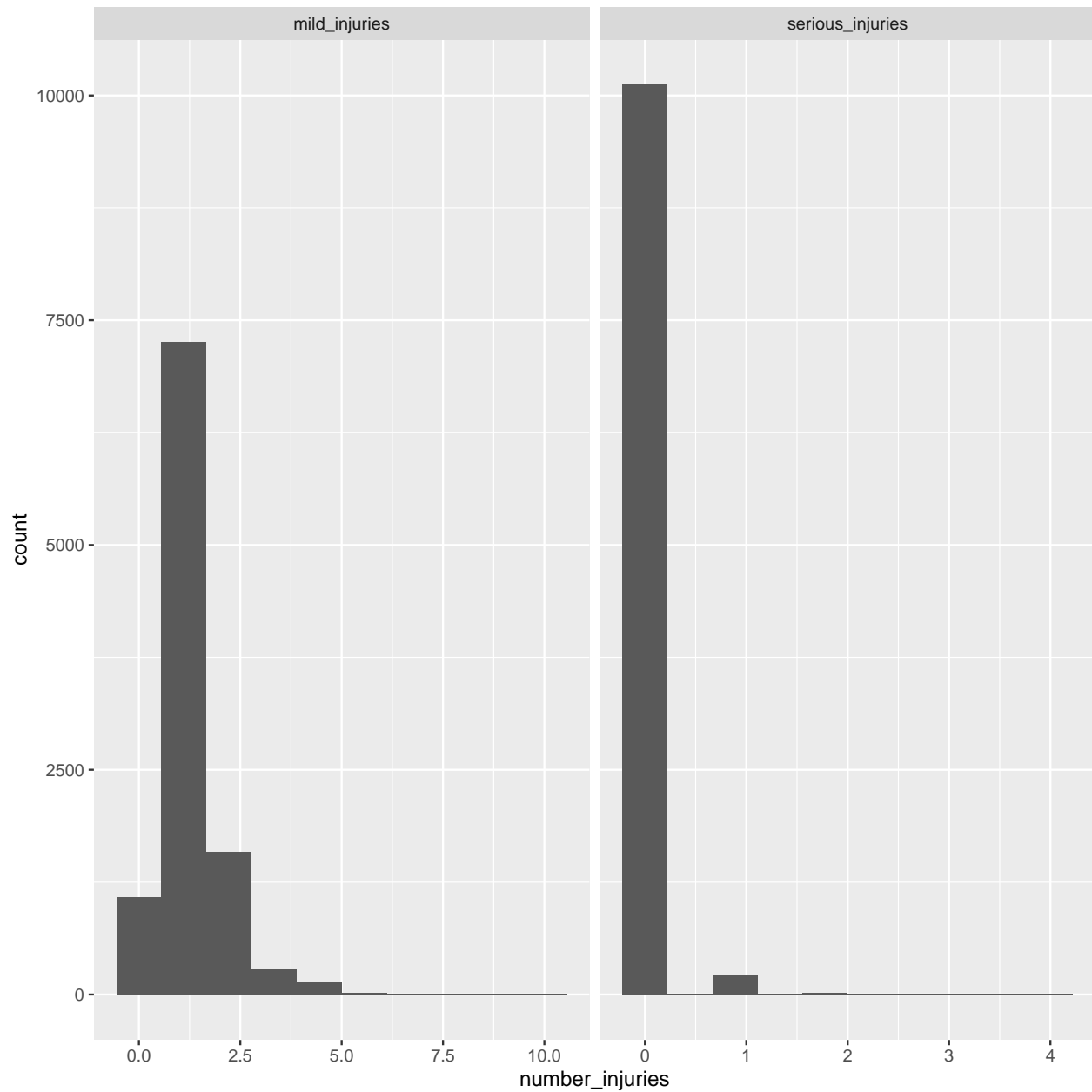
```
##
##
##
## [1] "transport"
##   code   transport longitude latitude
## 1: K001 Underground  2.119370 41.39920
## 2: K001 Underground  2.135427 41.39779
## 3: K001 Underground  2.185391 41.45149
## 4: K001 Underground  2.174473 41.46089
## 5: K001 Underground  2.168588 41.38720
## 6: K001 Underground  2.168507 41.38771
##
##                                     station
## 1: FGC (L6) - REINA ELISENDA (Sortida Duquesa d'Orleans)-
## 2:                                     FGC (L6) - LA BONANOVA-
## 3:                               METRO (L11) - CASA DE L'AIGUA (C. Vila-Real)-
## 4:                               METRO (L11) - CIUTAT MERIDIANA (C. Pedraforca)-
## 5:                               METRO (L1) - CATALUNYA (Rda. Universitat)-
## 6:                               METRO (L1) - CATALUNYA (Rambla de Catalunya)-
##   district.name      neighborhood.name
## 1: Sarrià-Sant Gervasi      Sarrià
## 2: Sarrià-Sant Gervasi Sant Gervasi - Galvany
## 3:      Nou Barris      la Trinitat Nova
## 4:      Nou Barris      Ciutat Meridiana
## 5:      Eixample la Dreta de l'Eixample
## 6:      Eixample la Dreta de l'Eixample
##
##   code      transport      longitude      latitude
## Length:651      Length:651      Min.      :2.005      Min.      :41.30
## Class :character      Class :character      1st Qu.:2.135      1st Qu.:41.38
## Mode  :character      Mode  :character      Median :2.164      Median :41.40
##                                     Mean  :2.157      Mean  :41.40
##                                     3rd Qu.:2.182      3rd Qu.:41.42
##                                     Max.   :2.249      Max.   :41.50
##   station      district.name      neighborhood.name
## Length:651      Length:651      Length:651
## Class :character      Class :character      Class :character
## Mode  :character      Mode  :character      Mode  :character
##
##
##
```

Accidents

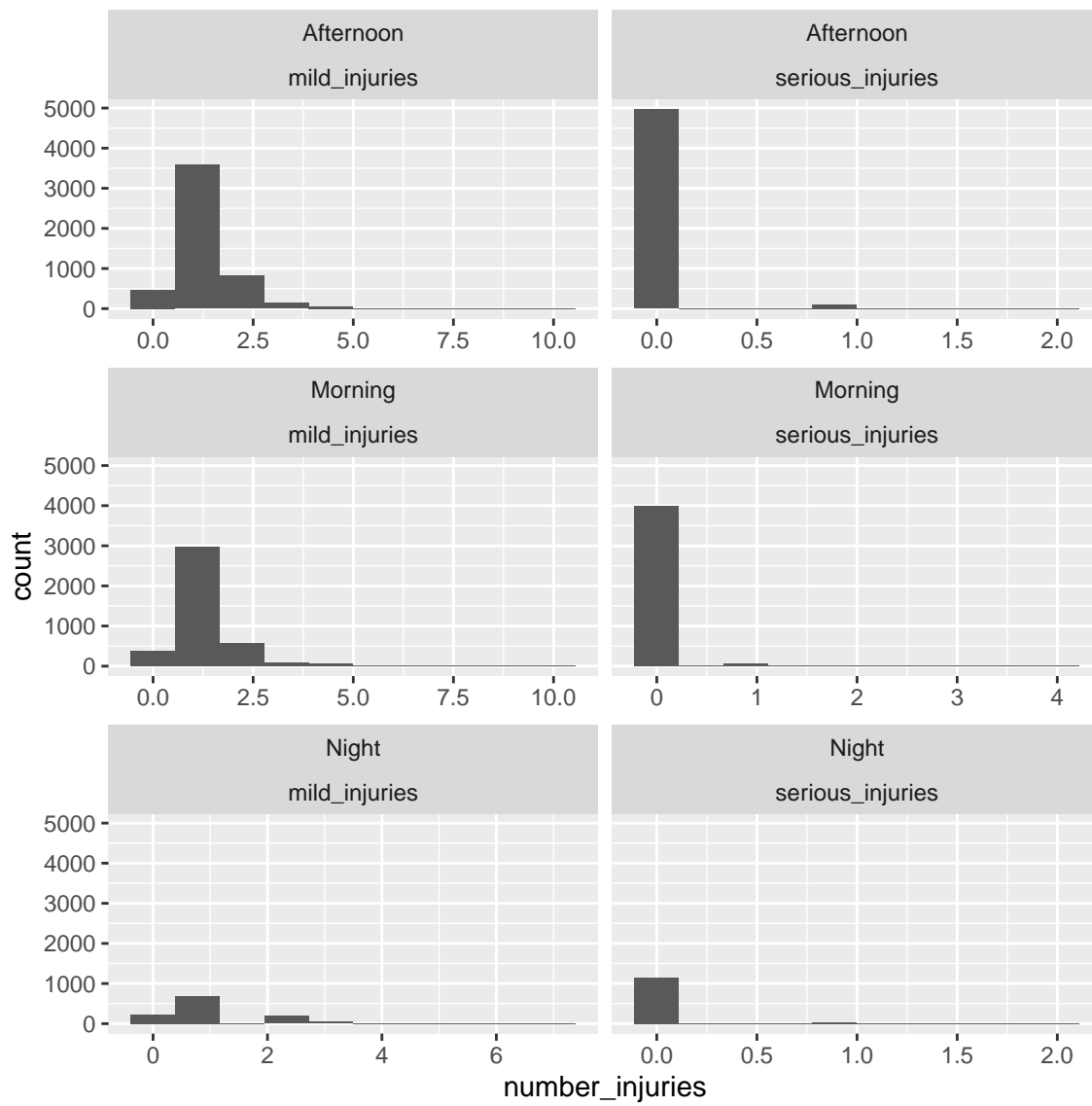
```
accidents[, dates := as.Date(paste0(month, "/", day, "/2017"), "%B/%d/%Y")]
# accidents[, district_name := as.factor(district_name)]
accidents_injuries <- melt(accidents,
                           measure.vars = c("mild_injuries", "serious_injuries"),
                           value.name = "number_injuries",
                           variable.name = "injury_type")

ggplot(accidents_injuries, aes(x=number_injuries)) +
  geom_histogram(bins=10) +
```

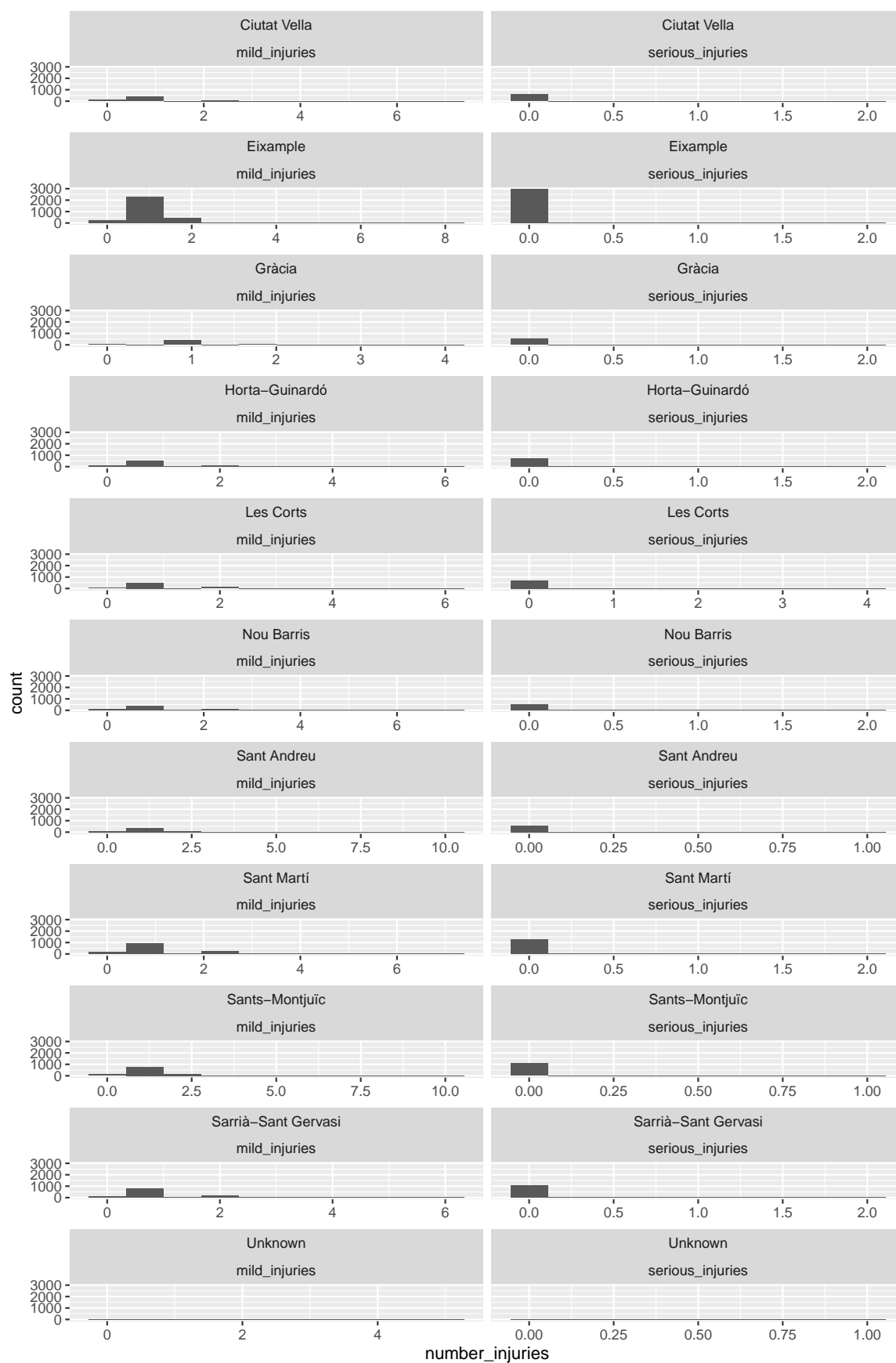
```
facet_wrap(~injury_type, ncol = 2, scales="free_x")
```



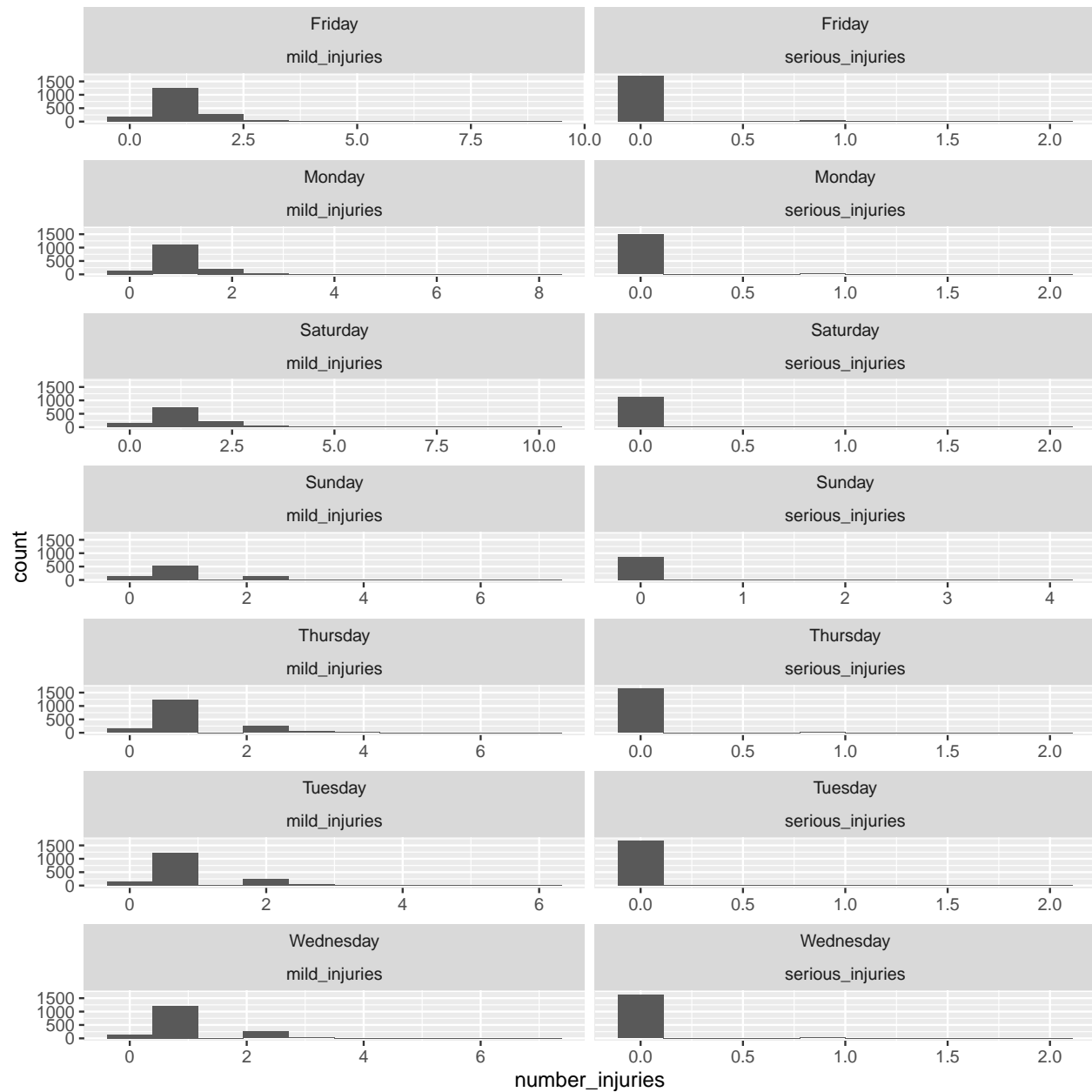
```
ggplot(accidents_injuries, aes(x=number_injuries)) +  
  geom_histogram(bins=10) +  
  facet_wrap(~part_of_the_day+injury_type, ncol = 2, scales="free_x")
```



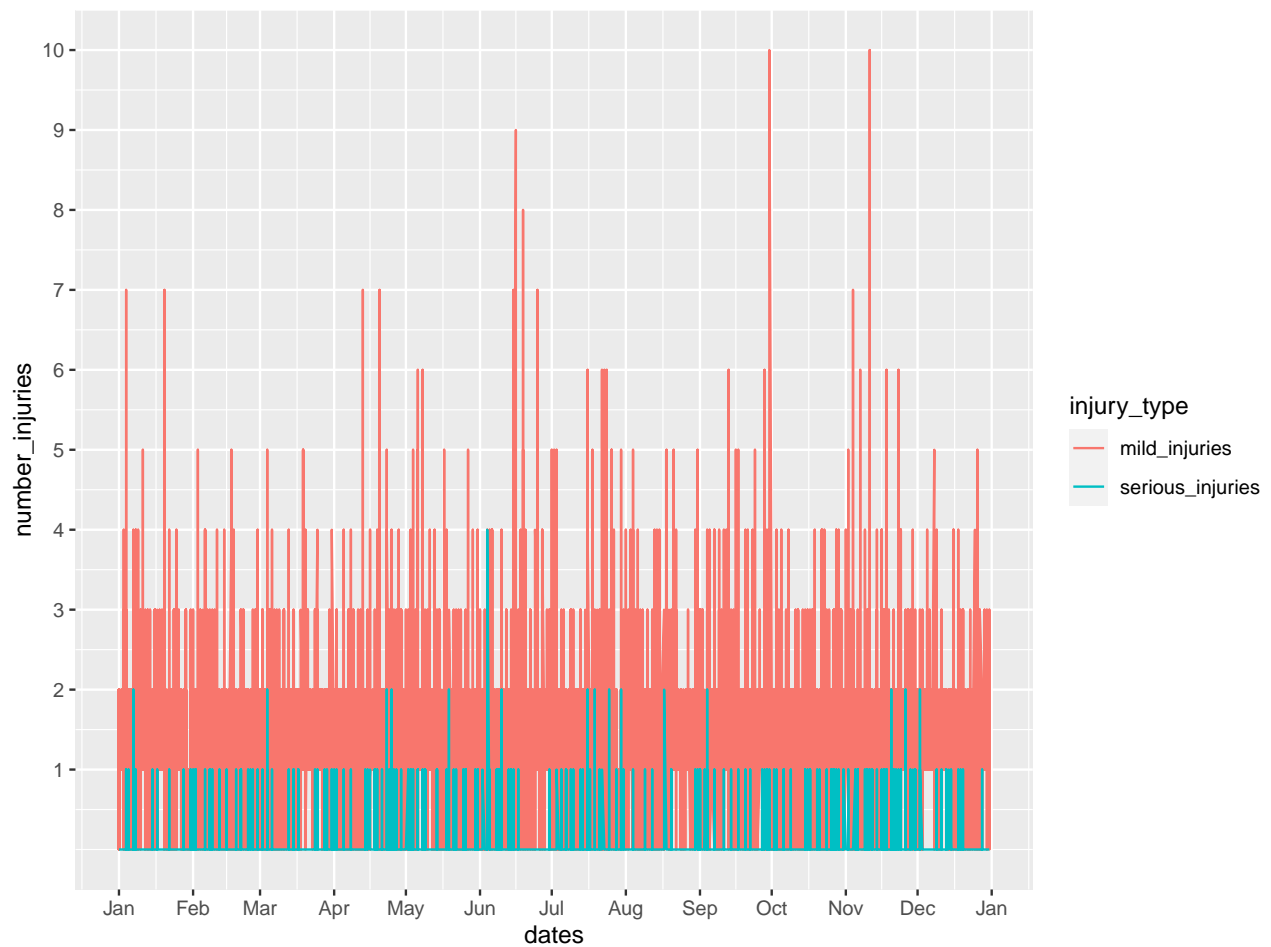
```
ggplot(accidents_injuries, aes(x=number_injuries)) +
  geom_histogram(bins=10) +
  facet_wrap(~district_name+injury_type, ncol = 2, scales="free_x")
```



```
ggplot(accidents_injuries, aes(x=number_injuries)) +
  geom_histogram(bins=10) +
  facet_wrap(~weekday+injury_type, ncol = 2, scales="free_x")
```



```
ggplot(accidents_injuries, aes(y=number_injuries, x=dates, color=injury_type)) +
  geom_line() +
  scale_x_date(date_breaks = "1 month", date_labels = "%b") +
  scale_y_continuous(breaks=1:10)
```



Bus Stops, Transport

```
bus_stops[, .N, by=transport]
```

```
##      transport      N
## 1:   Day bus stop 2324
## 2:  Night bus stop  828
## 3: Airport bus stop    7
## 4:   Bus station    3
```

```
bus_stops[, bus_prefix := strsplit(bus.stop, " ")[[1]][1], by=bus.stop]
bus_stops[, bus_per_station := strsplit(bus.stop, " ")[[1]][2], by=bus.stop]
bus_stops[, .(code, bus_prefix, transport)] %>% distinct()
```

```
##   code bus_prefix      transport
## 1: K014      BUS   Day bus stop
## 2: K015   NITBUS  Night bus stop
## 3: K016  AEROBUS Airport bus stop
## 4: K017  Estació   Bus station
```

```
bus_stops_per_bus <- as.data.table(separate_rows(bus_stops, 9, sep = "-"))[bus_per_station != ""]
bus_stops_per_bus[, c("bus.stop"):=NULL]
```



```
head(bus_stops_per_bus)
```

```
##      code      transport longitude latitude      district.name
## 1: K014 Day bus stop  2.171619 41.41374      Horta-Guinardó
## 2: K014 Day bus stop  2.134902 41.42022              Gràcia
## 3: K014 Day bus stop  2.162913 41.42319      Horta-Guinardó
## 4: K014 Day bus stop  2.163667 41.42290      Horta-Guinardó
## 5: K014 Day bus stop  2.120212 41.39721 Sarrià-Sant Gervasi
## 6: K014 Day bus stop  2.138941 41.41108 Sarrià-Sant Gervasi
##      neighborhood.name bus_prefix bus_per_station
## 1:              el Guinardó      BUS             192
## 2: Vallcarca i els Penitents      BUS             124
## 3:      la Font d'en Fargues      BUS             117
## 4:      la Font d'en Fargues      BUS             117
## 5:              Sarrià          BUS             130
## 6:      el Putxet i el Farró      BUS             131
```

```
summary(bus_stops_per_bus)
```

```
##      code      transport      longitude      latitude
## Length:5947      Length:5947      Min.      :2.056      Min.      :41.32
## Class :character Class :character 1st Qu.:2.142      1st Qu.:41.38
## Mode  :character Mode  :character Median :2.162      Median :41.40
##                                     Mean  :2.161      Mean  :41.40
##                                     3rd Qu.:2.180      3rd Qu.:41.42
##                                     Max.   :2.222      Max.   :41.47
## district.name      neighborhood.name      bus_prefix
## Length:5947      Length:5947      Length:5947
## Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character
##
##
##
## bus_per_station
## Length:5947
## Class :character
## Mode  :character
##
##
##
```

```
transport[, .N, by=transport]
```

```
##      transport      N
## 1:      Underground 463
## 2:      Railway (FGC) 54
## 3:      Tram        65
## 4: Maritime station 16
## 5:      RENFE       33
## 6:      Airport train 9
## 7:      Cableway    5
## 8:      Funicular    6
```

```
transport[, .N, by=code]
```

```
##      code      N
```

```
## 1: K001 463
## 2: K002 54
## 3: K011 65
## 4: K008 16
## 5: K003 33
## 6: K004 9
## 7: K010 5
## 8: K009 6
```

Explore districts

```
accidents[, victims_per_district:=sum(victims), by=district_name]
districts <- accidents[order(victims_per_district), unique(district_name)]
for (dist in districts){
  print(dist)
  print(paste0("Number of victims in total: ", accidents[district_name == dist, unique(victims_per_district)]))
  print(paste0("Number of victims with mild injuries: ", accidents[district_name == dist, sum(mild_injuries)]))
  print(paste0("Number of victims with serious injuries: ", accidents[district_name == dist, sum(serious_injuries)]))
  cat("\n")
  total_trans <- transport[district.name == dist, uniqueN(station), by=transport][, sum(V1)]
  print(paste0("Total different station-transport combinations: ", total_trans))
  print(transport[district.name == dist, uniqueN(station), by=transport])
  cat("\n")
  total_bus <- bus_stops_per_bus[district.name == dist, uniqueN(bus_per_station), by=transport][, sum(V1)]
  print(paste0("Total different busses: ", total_bus))
  print(bus_stops_per_bus[district.name == dist, uniqueN(bus_per_station), by=transport])
  cat("\n")
  print(paste0("Total of busses and station-transport combinations: ", total_bus + total_trans))
  cat("\n\n")
}
```

```
## [1] "Unknown"
## [1] "Number of victims in total: 40"
## [1] "Number of victims with mild injuries: 39"
## [1] "Number of victims with serious injuries: 1"
##
## [1] "Total different station-transport combinations: 0"
## Empty data.table (0 rows and 2 cols): transport,V1
##
## [1] "Total different busses: 0"
## Empty data.table (0 rows and 2 cols): transport,V1
##
## [1] "Total of busses and station-transport combinations: 0"
##
##
## [1] "Gràcia"
## [1] "Number of victims in total: 588"
## [1] "Number of victims with mild injuries: 576"
## [1] "Number of victims with serious injuries: 12"
##
## [1] "Total different station-transport combinations: 17"
##      transport V1
## 1:   Underground 15
```

```

## 2: Railway (FGC) 2
##
## [1] "Total different busses: 29"
##      transport V1
## 1:   Day bus stop 25
## 2: Night bus stop 4
##
## [1] "Total of busses and station-transport combinations: 46"
##
##
## [1] "Ciutat Vella"
## [1] "Number of victims in total: 634"
## [1] "Number of victims with mild injuries: 623"
## [1] "Number of victims with serious injuries: 11"
##
## [1] "Total different station-transport combinations: 27"
##      transport V1
## 1:   Underground 22
## 2:   Cableway    2
## 3:   RENFE       1
## 4: Maritime station 2
##
## [1] "Total different busses: 33"
##      transport V1
## 1:   Day bus stop 24
## 2: Night bus stop 9
##
## [1] "Total of busses and station-transport combinations: 60"
##
##
## [1] "Nou Barris"
## [1] "Number of victims in total: 661"
## [1] "Number of victims with mild injuries: 647"
## [1] "Number of victims with serious injuries: 13"
##
## [1] "Total different station-transport combinations: 37"
##      transport V1
## 1: Underground 36
## 2:   RENFE     1
##
## [1] "Total different busses: 31"
##      transport V1
## 1:   Day bus stop 27
## 2: Night bus stop 4
##
## [1] "Total of busses and station-transport combinations: 68"
##
##
## [1] "Sant Andreu"
## [1] "Number of victims in total: 704"
## [1] "Number of victims with mild injuries: 693"
## [1] "Number of victims with serious injuries: 11"
##
## [1] "Total different station-transport combinations: 34"

```

```

##      transport V1
## 1:   Underground 30
## 2:      RENFE    3
## 3: Airport train 1
##
## [1] "Total different busses: 38"
##      transport V1
## 1:   Day bus stop 31
## 2: Night bus stop 6
## 3:   Bus station 1
##
## [1] "Total of busses and station-transport combinations: 72"
##
##
## [1] "Horta-Guinardó"
## [1] "Number of victims in total: 839"
## [1] "Number of victims with mild injuries: 827"
## [1] "Number of victims with serious injuries: 10"
##
## [1] "Total different station-transport combinations: 29"
##      transport V1
## 1: Underground 29
##
## [1] "Total different busses: 37"
##      transport V1
## 1:   Day bus stop 32
## 2: Night bus stop 5
##
## [1] "Total of busses and station-transport combinations: 66"
##
##
## [1] "Les Corts"
## [1] "Number of victims in total: 894"
## [1] "Number of victims with mild injuries: 865"
## [1] "Number of victims with serious injuries: 27"
##
## [1] "Total different station-transport combinations: 21"
##      transport V1
## 1:      Tram    9
## 2: Underground 12
##
## [1] "Total different busses: 43"
##      transport V1
## 1:   Day bus stop 36
## 2: Night bus stop 7
##
## [1] "Total of busses and station-transport combinations: 64"
##
##
## [1] "Sarrià-Sant Gervasi"
## [1] "Number of victims in total: 1290"
## [1] "Number of victims with mild injuries: 1266"
## [1] "Number of victims with serious injuries: 24"
##

```

```

## [1] "Total different station-transport combinations: 44"
##      transport V1
## 1:  Underground 14
## 2: Railway (FGC) 17
## 3:      Tram    9
## 4:  Funicular   4
##
## [1] "Total different busses: 43"
##      transport V1
## 1:  Day bus stop 36
## 2: Night bus stop 7
##
## [1] "Total of busses and station-transport combinations: 87"
##
##
## [1] "Sants-Montjuïc"
## [1] "Number of victims in total: 1329"
## [1] "Number of victims with mild injuries: 1305"
## [1] "Number of victims with serious injuries: 21"
##
## [1] "Total different station-transport combinations: 50"
##      transport V1
## 1:  Underground 36
## 2: Maritime station 5
## 3:      Cableway 3
## 4:      Funicular 2
## 5:  Railway (FGC) 2
## 6:      RENFE    1
## 7:  Airport train 1
##
## [1] "Total different busses: 63"
##      transport V1
## 1:  Day bus stop 48
## 2:  Night bus stop 12
## 3:  Bus station 1
## 4: Airport bus stop 2
##
## [1] "Total of busses and station-transport combinations: 113"
##
##
## [1] "Sant Martí"
## [1] "Number of victims in total: 1645"
## [1] "Number of victims with mild injuries: 1593"
## [1] "Number of victims with serious injuries: 51"
##
## [1] "Total different station-transport combinations: 62"
##      transport V1
## 1:  Underground 44
## 2:      Tram    16
## 3: Airport train 1
## 4:      RENFE    1
##
## [1] "Total different busses: 43"
##      transport V1

```

```
## 1: Day bus stop 32
## 2: Night bus stop 11
##
## [1] "Total of busses and station-transport combinations: 105"
##
##
## [1] "Eixample"
## [1] "Number of victims in total: 3562"
## [1] "Number of victims with mild injuries: 3499"
## [1] "Number of victims with serious injuries: 60"
##
## [1] "Total different station-transport combinations: 106"
##      transport V1
## 1: Underground 94
## 2: Railway (FGC) 7
## 3: RENFE 3
## 4: Airport train 1
## 5: Tram 1
##
## [1] "Total different busses: 76"
##      transport V1
## 1: Night bus stop 19
## 2: Day bus stop 54
## 3: Airport bus stop 2
## 4: Bus station 1
##
## [1] "Total of busses and station-transport combinations: 182"
```

Conclusion: Total number of victims tends to be higher for increasing number of busses and transport opportunities per district.

But: could be both related to district size.

```
air_general <- merge(air_quality,
                    air_stations,
                    by="station",
                    all.x = TRUE)[, c("air_quality", "district_name")][, value := 1]
air_gen <- dcast(air_general,
                ... ~ air_quality,
                value.var = "value")
```

Aggregate function missing, defaulting to 'length'

```
air_gen

##      district_name -- Good Moderate
## 1:      <NA>      8    710         0
## 2: Ciutat Vella 17    701         0
## 3:      Eixample  0    635        83
## 4:      Gracia  66    630        22
## 5: Horta-Guinardo 2    716         0
## 6:      Les Corts  5    710         3
## 7:      Sant Marti 16   680        22
## 8: Sants-Montjuic 47   671         0
```

Outlier: district 2 is interesting :)