

Pmweb Report

Paulo Souza Junior

January 23, 2018

Introdução

Este documento apresenta um relatório analisando os dados abertos da cidade de Porto Alegre. O mesmo analisa dados relativos aos anos 2000 até 2016. Este relatório é gerado automaticamente através da linguagem R, pelo R Markdown e R Knit.

Inicialmente os arquivos a serem analisados são carregados e estruturados de forma equivalente para garantir uma análise combinatória destes dados. Isto é necessário, visto que os dados disponibilizados em [datapoa](http://datapoa.com.br)¹ não possuem a mesma distribuição de colunas e formato de datas.

```
path = "~/Documents/pmweb/dataset/"

fNames <- list.files(path, pattern = "acidentes-2*")

acidentes <- lapply(paste(path, fNames, sep=""), function(fNames){
  data.frame(read.csv(fNames, header=TRUE, sep=";"))
})

aux <- data.frame()
year <- 2000

for(i in 1:17){
  acidentes[[i]]$year <- year
  if(i > 15){
    acidentes[[i]]$DATA_HORA <- format(as.POSIXlt(acidentes[[i]]$DATA_HORA, format="%Y-%m-%dT%H:%M"), "%Y-%m-%dT%H:%M")
  }
  aux <- rbind(aux, select(acidentes[[i]], DATA_HORA, TEMPO, UPS, year))
  year = year + 1
}

aux$freq <- 1
```

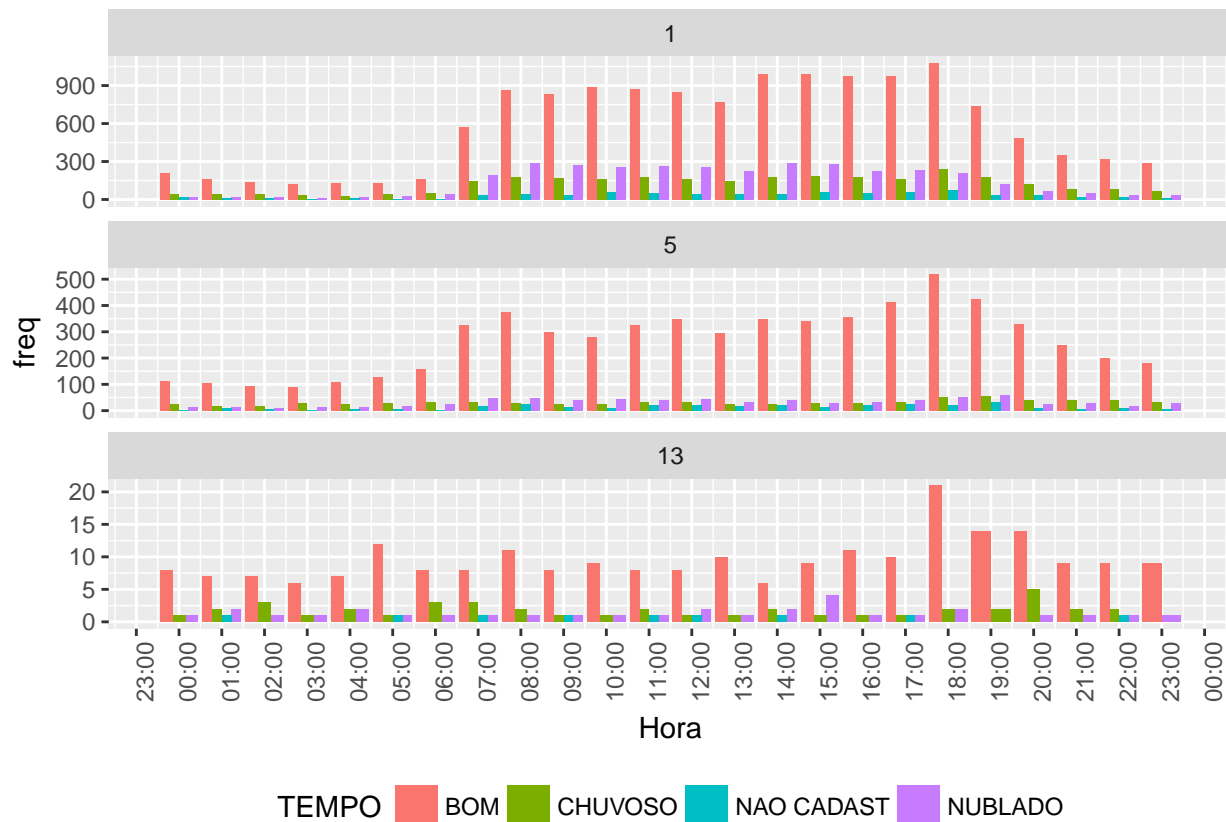
Condições atmosféricas, severidade e horário de acidentes

As relações das condições atmosféricas, severidade dos acidentes e horário das ocorrências são apresentadas nesta seção. O gráfico a seguir apresenta a frequência de ocorrência a cada hora, em um período de 24 horas. As severidades das ocorrências são classificadas em 1, 5 e 15, sendo (1 acidente com danos materiais, 5 acidente com ferido, 15 acidente com morte), estes são apresentados em um grid de gráficos. O clima é representado pela variável TEMPO, sendo eles BOM, CHUVOSO, NÃO CADASTRADO e NUBLADO e são representados pelas cores no gráfico.

```
plot <- aux %>% group_by(year, UPS, TEMPO, time =
  floor_date(as.POSIXct(format(strptime(DATA_HORA,
    "%Y%m%d %H:%M"), format="%H:%M"), format="%H:%M"), "1 hours")) %>%
  summarise(freq = n())
```

¹<http://datapoa.com.br>

```
ggplot(plot[complete.cases(plot),], aes(x = as.POSIXct(time, format="%H:%M"),
    y = freq, fill= TEMPO)) + geom_bar(stat = 'identity', position = 'dodge') +
labs(x="Hora") + facet_wrap(~UPS, nrow=3, scales="free_y") +
scale_x_datetime(date_breaks = "1 hour", date_labels = "%H:%M") +
theme(axis.text.x = element_text(angle = 90, hjust = 1), legend.position = "bottom")
```



É possível perceber que a maioria das ocorrências são sobre o tempo bom, tendo seu pico inicial a partir das 6 da manhã. A taxa de ocorrência reduz ao longo do dia, principalmente as 12 horas, tendo novamente um aumento as 18 horas. Logo após as ocorrências de tempo bom, as ocorrências de tempo nublado estão como parte da maioria, principalmente em casos de severidade 1. Em relação as severidades: casos de severidade 1 são mais frequentes sendo aproximadamente 42% dos casos, seguidos dos casos 5 e 13, 38% e 19% respectivamente. Nos casos de severidade 5, o comportamento das ocorrências não varia em relação ao clima, o total de ocorrências para o tempo chuvoso e nublado é muito próximo. Já para os casos de severidade 13 o clima chuvoso é significativo, comparado com as severidades anteriores, onde predomina em relação ao clima nublado e não cadastrado. O mesmo é significativamente alto em horários como 2h, 4h, 6h, 7h da manhã e durante a noite das 18h e 20h. É possível apontar também, que para horários como as 18 horas, que possui o maior índices de ocorrências severas, não é frequente em clima chuvoso.

Além disso, é interessante comentar que a opção para clima NAO CADAST (não cadastrado), não foi utilizado em alguns anos ou foi utilizado muito pouco. Em anos como 2000, 2002~2007 quase não houve casos neste fator e em alguns anos altos índices estão presentes (e.g. 2001 e 2009). Como pode ser visto no gráfico abaixo, que apresenta fator de severidade por horário e clima em cores/formas.

```
ggplot(plot[complete.cases(plot),], aes(x = time,
    y = UPS, shape= TEMPO, colour=TEMPO)) + geom_jitter() +
facet_wrap(~year, ncol=3) +
ggtitle("Total de acidentes x Hora x Clima") +
scale_x_datetime(date_breaks = "5 hour", date_labels = "%H:%M") +
```

```
theme(axis.text.x = element_text(angle = 90, hjust = 1), legend.position = "bottom")
```

Total de acidentes x Hora x Clima



Danos causados em acidentes

Nesta seção é apresentado os veículos que mais causaram acidentes. Também

Foram encontrados problemas na linha 11084 do dataset relativo ao ano de 2014, que resultou em valores incorretos para a leitura dos acidentes.

```
perdas <- data.frame()

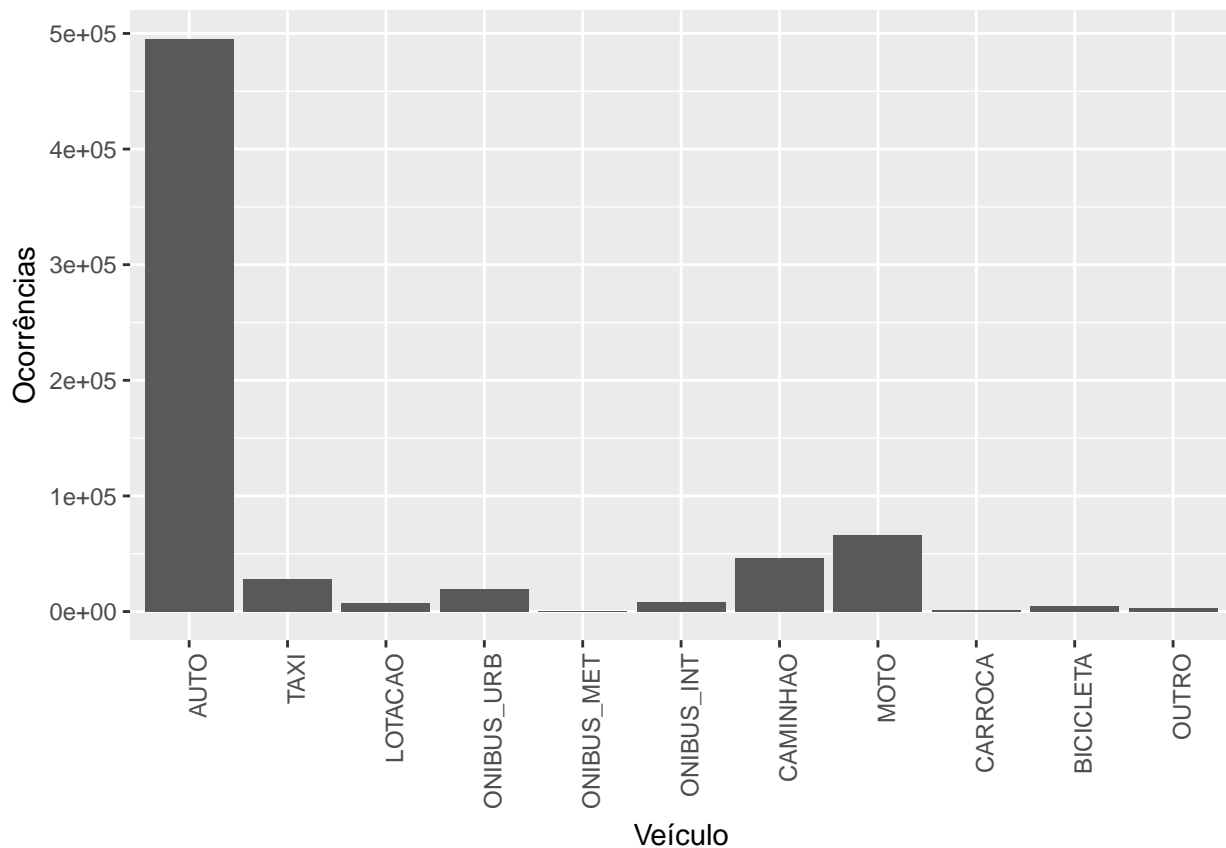
for(i in 1:17){
  acidentes[[i]]$year <- year
  if(i < 15){
    acidentes[[i]]$ONIBUS_MET <- 0
    acidentes[[i]]$FERIDOS_GR <- 0
  }
  perdas <- rbind(perdas, select(acidentes[[i]], ID, AUTO, TAXI, LOTACAO,
                                ONIBUS_URB, ONIBUS_MET, ONIBUS_INT, CAMINHAO,
                                MOTO, CARROCA, BICICLETA, OUTRO, FERIDOS, FERIDOS_GR))

  year = year + 1
}

perdas <- as.data.frame(lapply(perdas, function(x) as.numeric(as.character(x))))
materiais <- perdas %>% summarize_all(funs(sum(., na.rm=TRUE))) %>% melt()
```

No id variables; using all as measure variables

```
ggplot(materiais[2:12,], aes(x = variable,
                             y = value)) + geom_bar(stat = 'identity', position = 'dodge') +
  labs(x="Veículo", y = "Ocorrências") + theme(axis.text.x = element_text(angle = 90, hjust = 1), legend = FALSE)
```



```
ggplot(materiais[3:12,], aes(x = variable,
                             y = value)) + geom_bar(stat = 'identity', position = 'dodge') +
```

```
labs(x="Veículo", y = "Ocorrências") + theme(axis.text.x = element_text(angle = 90, hjust = 1), legend
```

