reddit submissions analysis

Daily Reddit Submissions from 2006 - 2015

This assumes that all packages in "Rstart.R" are installed.

```
options(warn=-1)
source("Rstart.R")
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
## Registering fonts with R
```

```
##
## Attaching package: 'scales'
```

```
## The following objects are masked from 'package:readr':
##
##     col_factor, col_numeric
```

```
library(tidyr)
library(bigrquery)
library(wordcloud)
library(digest)

options(repr.plot.mimetypes = 'image/png', repr.plot.width=4, repr.plot.height=3, repr.plot.res=300)
project_name <- "long-micron-152118"

sql <- "SELECT DATE(SEC_TO_TIMESTAMP(created)) date_submission,
COUNT(*) as num_submissions
FROM [fh-bigquery:reddit_posts.full_corpus_201509]
GROUP BY date_submission
ORDER by date_submission"

df <- tbl_df(query_exec(sql, project=project_name, max_pages=Inf))
df %>% tail(10)
```

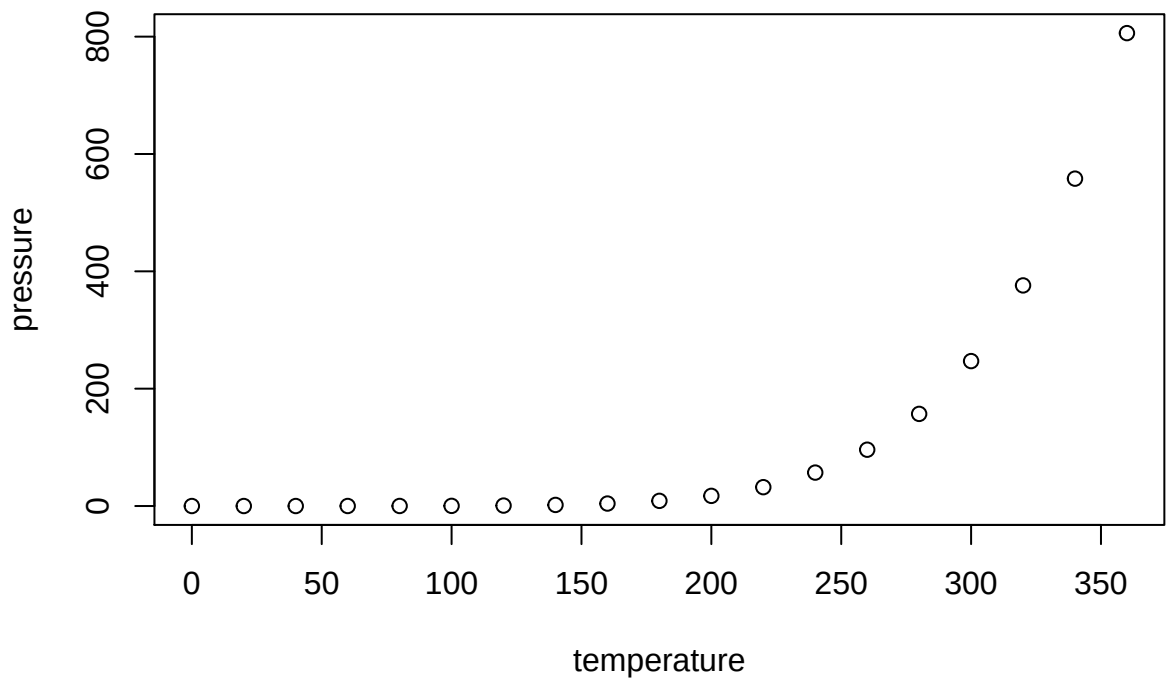
```
## # A tibble: 10 × 2
##    date_submission num_submissions
##              <chr>           <int>
## 1       2015-08-23          170999
## 2       2015-08-24          163107
## 3       2015-08-25          264787
## 4       2015-08-26          235858
## 5       2015-08-27          212472
## 6       2015-08-28          206100
## 7       2015-08-29          180039
```

```
## 8       2015-08-30          183686
## 9       2015-08-31          214685
## 10      2015-09-01           10299
```

```
plot <- ggplot(df, aes(x=as.Date(date_submission), y=num_submissions)) +
  geom_area(fill= "purple", alpha=0.85, size=0) + #"#2980b8"
  fte_theme() +
  ylim(0,245000)+
  scale_x_date(breaks=date_breaks("1 year"), labels=date_format("%Y")) +
  labs(x="Year", y="Submissions", title="Daily Reddit Submissions from 2006 - 2015")
#plot
max_save(plot, "reddit-bigquery-1", "Reddit")
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.