Bicycle Rentals

Data Computing

Computing project

In this activity, you'll examine some factors that may influence the use of bicycles in a bike-renting program. The data come from Washington, DC and cover the last quarter of 2014.

We will use the "Trips-history" data file You can access the data like this.

```r
data_site <- "http://tiny.cc/dcf/2014-Q4-Trips-History-Data-Small.rds"
Trips <- readRDS(gzcon(url(data_site)))
```

Important: To avoid repeatedly re-reading the files, start the above chunk with {r cache = TRUE} rather than the usual {r}.

The `Trips` data table is a random subset of 10,000 trips from the full quarterly data. Start with this small data table to develop your analysis commands. When you have this working well, you can access the full data set of more than 600,000 events by removing `-Small` from the name of the data_site.

Time of day

It's natural to expect that bikes are rented more at some times of day than others. The variable `sdate` gives the time (including the date) that the rental started.

Make these plots and interpret them:

1. A density plot of the events versus `sdate`. Use `ggplot()` and `geom_density()`.

2. A density plot of the events versus time of day. You can use `lubridate::hour()`, and `lubridate::minute()` to extract the hour of the day and minute within the hour from `sdate`, e.g.

   ```r
   Trips %>%
     mutate(time_of_day =
           lubridate::hour(sdate) +
             lubridate::minute(sdate) / 60) %>%
     ... further processing ...
   ```

3. Facet (2) by day of the week. (Use `lubridate::wday()` to generate day of the week. Use `+ facet_wrap( ~ day_of_week)` to perform the faceting.)

4. Set the `fill` aesthetic for `geom_density()` to the `client` variable.[1] You may also want to set the `alpha` for transparency and `color=NA` to suppress the outline of the density function.

5. Same as (4) but using `geom_density()` with the argument `position = position_stack()`.

- In your opinion, which of these graphics is most effective at telling an interesting story?

6. Rather than faceting on day of the week, consider creating a new faceting variable like this:

```
mutate(wday = ifelse(lubridate::wday(sdate) %in% c(1,7), "weekend", "weekday"))
```

- Is it better to facet on `wday` and fill with `client`, or vice versa?

[1] `client` describes whether the renter is a regular user (level `Registered`) or has not joined the bike-rental organization (`Causal`).