Homework 04

DSO 545: Statistical Computing and Data Visualization

Spring 2017

Due Date: Monday, March 5, 2018 (midnight)

Boston Bike Rides

When do people in Boston bike most? This time we are exploring the `Boston Hubway` dataset. It contains data about every trip taken in 2012 – with date, time, origin and destination stations, plus the bike number and more. Here is more information about the variables:

- **seq_id**: unique record ID
- **hubway_id**: trip id
- **status**: trip status; "closed" indicates a trip has terminated
- **duration**: time of trip in seconds
- **start_date**: start date of trip with date and time, in EST (hint: be careful with the time zone!)
- **strt_statn**: id of start station
- **end_date**: end date of trip with date and time, in EST (hint: be careful with the time zone!)
- **end_statn**: station id of end station
- **bike_nr**: id of bicycle used
- **subsc_type**: subscription type - "Registered" is user with membership; "Casual" is user without membership
- **zip_code**: zipcode of user (only available for registered users)
- **birth_date**: birth year of user
- **gender**: gender of user

The dataset is stored in an RDA file ("hubwaytrip2012.rda").

Questions

1. Use functions from `dplyr` and `lubridate` to summarise by `Day` and `Start Time (Hour of the day)` of bike rental usage. Save your new dataframe in a variable called `trips2012Time`, and output `head(trips2012Time)` to show the first few observations of the new dataframe. (Make sure to set the timezone to "EST")

2. Create the following heatmap , use color `#e5f5e0` and `#31a354`. **Summarise what you see**.