

Update the data

Georgie Knight

20 August, 2016

We'll update our data to take account of the following:

Station 23: From 9/1/14 - 10/22/14: This station was located at (37.488501, -122.231061).

Station 25: From 9/1/14 - 10/22/14: This station was located at (37.486725, -122.225551). It was previously named "Broadway at Main."

Station 49: From 9/1/14 - 2/5/15: This station was located at (37.789625, -122.390264).

Station 69: From 9/1/14 - 3/11/15: This station was located at (37.776377, -122.39607).

Station 72: Moved twice. From 9/1/14 - 2/12/15, this station was located at (37.780356, -122.412919). From 2/13/15 to 6/3/15, the station was located at (37.780353, -122.41226).

Station 80: On 9/1/14, this station changed names from "San Jose Government Center" to "Santa Clara County Civic Center." It did not move.

Dplyr and tidyr

Load the packages and the data:

```
library("dplyr")
```

```
##
## 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
```

```
library("tidyr")
library("lubridate")
```

```
##
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':
##
##   date
```

```
trip_read    <- read.csv("trip_full.csv")
status_read  <- read.csv("status_full.csv")
trip_df      <- data.frame(trip_read)
status_df    <- data.frame(status_read)
trip         <- dplyr::tbl_df(trip_df)
status       <- dplyr::tbl_df(status_df)
```

Firstly the trip data. We'll get the start/end dates/times together into one column using *unite*

```
trip <- trip %>%
  unite(Start_trip, Start_Date, Start_time, sep= " ", remove = TRUE) %>%
  unite(End_trip, End_Date, End_time, sep= " ", remove = TRUE)
```

We'll now convert all the relevant data columns into a time format with the correct time zone.

```
trip <- trip %>%
  mutate(Start_trip = ymd_hm(Start_trip, tz = "America/Los_Angeles")) %>%
  mutate(End_trip   = ymd_hm(End_trip,   tz = "America/Los_Angeles")) %>%
  mutate(Date = as.Date(Date))
```

Now update the latitudes and longitudes for the start

```
trip <- trip %>%
  mutate(
    start_lat=ifelse(between(Date, as.Date("2014-09-01"), as.Date("2014-10-22"))
      & Start.Terminal ==23,37.488501, start_lat)
  ) %>%
  mutate(
    start_long=ifelse(between(Date, as.Date("2014-09-01"),as.Date("2014-10-22"))
      & Start.Terminal ==23,-122.231061, start_long)
  ) %>%
  mutate(
    start_lat=ifelse(between(Date, as.Date("2014-09-01"), as.Date("2014-10-22"))
      & Start.Terminal ==25,37.486725, start_lat)
  ) %>%
  mutate(
    start_long=ifelse(between(Date, as.Date("2014-09-01"),as.Date("2014-10-22"))
      & Start.Terminal ==25,-122.225551, start_long)
  ) %>%
  mutate(
    start_lat=ifelse(between(Date, as.Date("2014-09-01"), as.Date("2015-02-05"))
      & Start.Terminal ==49,37.789625, start_lat)
  ) %>%
  mutate(
    start_long=ifelse(between(Date, as.Date("2014-09-01"),as.Date("2015-02-05"))
      & Start.Terminal ==49,-122.390264, start_long)
  ) %>%
  mutate(
    start_lat=ifelse(between(Date, as.Date("2014-09-01"), as.Date("2015-03-11"))
      & Start.Terminal ==69,37.776377, start_lat)
  ) %>%
  mutate(
```

```

start_long=ifelse(between(Date, as.Date("2014-09-01"),as.Date("2015-03-11"))
                  & Start.Terminal ==69,-122.39607, start_long)
) %>%
mutate(
start_lat=ifelse(between(Date, as.Date("2014-09-01"), as.Date("2015-02-12"))
                  & Start.Terminal ==72,37.780356, start_lat)
) %>%
mutate(
start_long=ifelse(between(Date, as.Date("2014-09-01"),as.Date("2015-02-12"))
                  & Start.Terminal ==72,-122.412919, start_long)
) %>%
mutate(
start_lat=ifelse(between(Date, as.Date("2015-02-13"), as.Date("2015-06-03"))
                  & Start.Terminal ==72,37.780353, start_lat)
) %>%
mutate(
start_long=ifelse(between(Date, as.Date("2015-02-13"),as.Date("2015-06-03"))
                  & Start.Terminal ==72,-122.41226, start_long)
)

```

Let's check that this has worked!

```

check <-select(trip, Date, Start.Terminal, start_lat, start_long)
check <-filter(check, Start.Terminal ==69)
filter(check, Date > as.Date("2015-03-11"))

```

```

## # A tibble: 11,650 x 4
##       Date Start.Terminal start_lat start_long
##       <date>         <int>      <dbl>      <dbl>
## 1 2015-08-31           69  37.7766 -122.3955
## 2 2015-08-31           69  37.7766 -122.3955
## 3 2015-08-31           69  37.7766 -122.3955
## 4 2015-08-31           69  37.7766 -122.3955
## 5 2015-08-31           69  37.7766 -122.3955
## 6 2015-08-31           69  37.7766 -122.3955
## 7 2015-08-31           69  37.7766 -122.3955
## 8 2015-08-31           69  37.7766 -122.3955
## 9 2015-08-31           69  37.7766 -122.3955
## 10 2015-08-31          69  37.7766 -122.3955
## # ... with 11,640 more rows

```

```

filter(check, Date < as.Date("2015-03-11"))

```

```

## # A tibble: 10,052 x 4
##       Date Start.Terminal start_lat start_long
##       <date>         <int>      <dbl>      <dbl>
## 1 2015-03-10           69  37.77638 -122.3961
## 2 2015-03-10           69  37.77638 -122.3961
## 3 2015-03-10           69  37.77638 -122.3961
## 4 2015-03-10           69  37.77638 -122.3961
## 5 2015-03-10           69  37.77638 -122.3961
## 6 2015-03-10           69  37.77638 -122.3961

```

```
## 7 2015-03-10          69 37.77638 -122.3961
## 8 2015-03-10          69 37.77638 -122.3961
## 9 2015-03-10          69 37.77638 -122.3961
## 10 2015-03-10         69 37.77638 -122.3961
## # ... with 10,042 more rows
```

Now update the latitudes and longitudes for the end

```
trip <- trip %>%
  mutate(
    end_lat=ifelse(between(Date, as.Date("2014-09-01"), as.Date("2014-10-22"))
      & End.Terminal ==23,37.488501, end_lat)
  ) %>%
  mutate(
    end_long=ifelse(between(Date, as.Date("2014-09-01"),as.Date("2014-10-22"))
      & End.Terminal ==23,-122.231061, end_long)
  ) %>%
  mutate(
    end_lat=ifelse(between(Date, as.Date("2014-09-01"), as.Date("2014-10-22"))
      & End.Terminal ==25,37.486725, end_lat)
  ) %>%
  mutate(
    end_long=ifelse(between(Date, as.Date("2014-09-01"),as.Date("2014-10-22"))
      & End.Terminal ==25,-122.225551, end_long)
  ) %>%
  mutate(
    end_lat=ifelse(between(Date, as.Date("2014-09-01"), as.Date("2015-02-05"))
      & End.Terminal ==49,37.789625, end_lat)
  ) %>%
  mutate(
    end_long=ifelse(between(Date, as.Date("2014-09-01"),as.Date("2015-02-05"))
      & End.Terminal ==49,-122.390264, end_long)
  ) %>%
  mutate(
    end_lat=ifelse(between(Date, as.Date("2014-09-01"), as.Date("2015-03-11"))
      & End.Terminal ==69,37.776377, end_lat)
  ) %>%
  mutate(
    end_long=ifelse(between(Date, as.Date("2014-09-01"),as.Date("2015-03-11"))
      & End.Terminal ==69,-122.39607, end_long)
  ) %>%
  mutate(
    end_lat=ifelse(between(Date, as.Date("2014-09-01"), as.Date("2015-02-12"))
      & End.Terminal ==72,37.780356, end_lat)
  ) %>%
  mutate(
    end_long=ifelse(between(Date, as.Date("2014-09-01"),as.Date("2015-02-12"))
      & End.Terminal ==72,-122.412919, end_long)
  ) %>%
  mutate(
    end_lat=ifelse(between(Date, as.Date("2015-02-13"), as.Date("2015-06-03"))
      & Start.Terminal ==72,37.780353, end_lat)
  ) %>%
  mutate(
```

```
end_long=ifelse(between(Date, as.Date("2015-02-13"),as.Date("2015-06-03"))
                & Start.Terminal ==72,-122.41226, end_long)
)
```

Now for the status data

```
status <- mutate(status, Date = as.Date(Date))
```

```
status <- status %>%
  mutate(
    lat=ifelse(between(Date, as.Date("2014-09-01"), as.Date("2014-10-22"))
              & station_id == 23, 37.488501, lat)
  ) %>%
  mutate(
    long=ifelse(between(Date, as.Date("2014-09-01"),as.Date("2014-10-22"))
              & station_id == 23, -122.231061, long)
  ) %>%
  mutate(
    lat=ifelse(between(Date, as.Date("2014-09-01"), as.Date("2014-10-22"))
              & station_id == 25, 37.486725, lat)
  ) %>%
  mutate(
    long=ifelse(between(Date, as.Date("2014-09-01"),as.Date("2014-10-22"))
              & station_id == 25, -122.225551, long)
  ) %>%
  mutate(
    lat=ifelse(between(Date, as.Date("2014-09-01"), as.Date("2015-02-05"))
              & station_id == 49, 37.789625, lat)
  ) %>%
  mutate(
    long=ifelse(between(Date, as.Date("2014-09-01"),as.Date("2015-02-05"))
              & station_id == 49, -122.390264, long)
  ) %>%
  mutate(
    lat=ifelse(between(Date, as.Date("2014-09-01"), as.Date("2015-03-11"))
              & station_id == 69, 37.776377, lat)
  ) %>%
  mutate(
    long=ifelse(between(Date, as.Date("2014-09-01"),as.Date("2015-03-11"))
              & station_id == 69, -122.39607, long)
  ) %>%
  mutate(
    lat=ifelse(between(Date, as.Date("2014-09-01"), as.Date("2015-02-12"))
              & station_id == 72, 37.780356, lat)
  ) %>%
  mutate(
    long=ifelse(between(Date, as.Date("2014-09-01"),as.Date("2015-02-12"))
              & station_id == 72, -122.412919, long)
  ) %>%
  mutate(
    lat=ifelse(between(Date, as.Date("2015-02-13"), as.Date("2015-06-03"))
              & station_id == 72, 37.780353, lat)
  ) %>%
```

```
mutate(  
  long=ifelse(between(Date, as.Date("2015-02-13"),as.Date("2015-06-03"))  
    & station_id == 72, -122.41226, long)  
)
```

```
write.csv(status, file="status_full_updated.csv")
```

```
write.csv(trip, file="trip_full_updated.csv")
```