Look at the BABS data

Georgie Knight

15 August, 2016

We're going to take a look at the bay area bike share data. We have two years worth of data

Dplyr and tidyr

Load the *dplyr* and *tidyr* packages which will help us wrangle 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")
```

Loading up the data

Load the various files for year 1 part 1 (29/8/13 - 28/2/14) and year 1 part 2 (3/1/14 - 8/31/14) and year 2 (1/9/14 - 31-8/15) into R:

```
station_data_original_year1_part1 <- read.csv("C:/Users/Georgie/Desktop/BABS_Data/201402_station_data.c
status_data_original_year1_part1  <- read.csv("C:/Users/Georgie/Desktop/BABS_Data/201402_status_data.cs
trip_data_original_year1_part1    <- read.csv("C:/Users/Georgie/Desktop/BABS_Data/201402_trip_data.csv")
weather_data_original_year1_part1 <- read.csv("C:/Users/Georgie/Desktop/BABS_Data/201402_weather_data.c

station_data_original_year1_part2 <- read.csv("C:/Users/Georgie/Desktop/BABS_Data/201408_station_data.c
status_data_original_year1_part2  <- read.csv("C:/Users/Georgie/Desktop/BABS_Data/201408_status_data.cs
trip_data_original_year1_part2    <- read.csv("C:/Users/Georgie/Desktop/BABS_Data/201408_trip_data.csv")
weather_data_original_year1_part2 <- read.csv("C:/Users/Georgie/Desktop/BABS_Data/201408_weather_data.c

station_data_original_year2       <- read.csv("C:/Users/Georgie/Desktop/BABS_Data/201508_station_data.c
status_data_original_year2        <- read.csv("C:/Users/Georgie/Desktop/BABS_Data/201508_status_data.cs
trip_data_original_year2          <- read.csv("C:/Users/Georgie/Desktop/BABS_Data/201508_trip_data.csv")
weather_data_original_year2       <- read.csv("C:/Users/Georgie/Desktop/BABS_Data/201508_weather_data.c
```

Convert them to data frames:

```
station_data_frame_year1_part1  <- data.frame(station_data_original_year1_part1)
status_data_frame_year1_part1   <- data.frame(status_data_original_year1_part1)
trip_data_frame_year1_part1     <- data.frame(trip_data_original_year1_part1)
weather_data_frame_year1_part1  <- data.frame(weather_data_original_year1_part1)

station_data_frame_year1_part2  <- data.frame(station_data_original_year1_part2)
status_data_frame_year1_part2   <- data.frame(status_data_original_year1_part2)
trip_data_frame_year1_part2     <- data.frame(trip_data_original_year1_part2)
weather_data_frame_year1_part2  <- data.frame(weather_data_original_year1_part2)

station_data_frame_year2        <- data.frame(station_data_original_year2)
status_data_frame_year2         <- data.frame(status_data_original_year2)
trip_data_frame_year2           <- data.frame(trip_data_original_year2)
weather_data_frame_year2        <- data.frame(weather_data_original_year2)
```

Convert them all to tables within the dplyr package...

```
station_data_table_year1_part1  <- dplyr::tbl_df(station_data_original_year1_part1)
status_data_table_year1_part1   <- dplyr::tbl_df(status_data_original_year1_part1)
trip_data_table_year1_part1     <- dplyr::tbl_df(trip_data_original_year1_part1)
weather_data_table_year1_part1  <- dplyr::tbl_df(weather_data_original_year1_part1)

station_data_table_year1_part2  <- dplyr::tbl_df(station_data_original_year1_part2)
status_data_table_year1_part2   <- dplyr::tbl_df(status_data_original_year1_part2)
trip_data_table_year1_part2     <- dplyr::tbl_df(trip_data_original_year1_part2)
weather_data_table_year1_part2  <- dplyr::tbl_df(weather_data_original_year1_part2)

station_data_table_year2 <- dplyr::tbl_df(station_data_original_year2)
status_data_table_year2  <- dplyr::tbl_df(status_data_original_year2)
trip_data_table_year2    <- dplyr::tbl_df(trip_data_original_year2)
weather_data_table_year2 <- dplyr::tbl_df(weather_data_original_year2)
```

Take a quick look at them:

```
## Observations: 69
## Variables: 7
## $ station_id   <int> 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 2...
## $ name         <fctr> San Jose Diridon Caltrain Station, San Jose Civi...
## $ lat          <dbl> 37.32973, 37.33070, 37.33399, 37.33141, 37.33672,...
## $ long         <dbl> -121.9018, -121.8890, -121.8949, -121.8932, -121....
## $ dockcount    <int> 27, 15, 11, 19, 15, 15, 15, 15, 15, 19, 19, 15, 1...
## $ landmark     <fctr> San Jose, San Jose, San Jose, San Jose, San Jose...
## $ installation <fctr> 8/6/2013, 8/5/2013, 8/6/2013, 8/5/2013, 8/7/2013...


## Observations: 16,994,602
## Variables: 4
## $ station_id      <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2...
## $ bikes_available <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2...
## $ docks_available <int> 25, 25, 25, 25, 25, 25, 25, 25, 25, 25, 25, 25...
## $ time            <fctr> 2013/08/29 12:06:01, 2013/08/29 12:07:01, 201...


## Observations: 144,015
```

```
## Variables: 11
## $ Trip.ID            <int> 4576, 4607, 4130, 4251, 4299, 4927, 4500, 45...
## $ Duration           <int> 63, 70, 71, 77, 83, 103, 109, 111, 113, 114,...
## $ Start.Date         <fctr> 8/29/2013 14:13, 8/29/2013 14:42, 8/29/2013...
## $ Start.Station      <fctr> South Van Ness at Market, San Jose City Hal...
## $ Start.Terminal     <int> 66, 10, 27, 10, 66, 59, 4, 8, 66, 10, 49, 6,...
## $ End.Date           <fctr> 8/29/2013 14:14, 8/29/2013 14:43, 8/29/2013...
## $ End.Station        <fctr> South Van Ness at Market, San Jose City Hal...
## $ End.Terminal       <int> 66, 10, 27, 10, 67, 59, 5, 8, 66, 11, 54, 4,...
## $ Bike..             <int> 520, 661, 48, 26, 319, 527, 679, 687, 553, 1...
## $ Subscription.Type  <fctr> Subscriber, Subscriber, Subscriber, Subscri...
## $ Zip.Code           <fctr> 94127, 95138, 97214, 95060, 94103, 94109, 9...


## Observations: 920
## Variables: 24
## $ Date                      <fctr> 8/29/2013, 8/30/2013, 8/31/2013, 9...
## $ Max_Temperature_F         <int> 74, 78, 71, 74, 75, 73, 74, 72, 85,...
## $ Mean_Temperature_F        <int> 68, 69, 64, 66, 69, 67, 68, 66, 71,...
## $ Min_TemperatureF          <int> 61, 60, 57, 58, 62, 60, 61, 60, 56,...
## $ Max_Dew_Point_F           <int> 61, 61, 57, 60, 61, 59, 59, 57, 57,...
## $ MeanDew_Point_F           <int> 58, 58, 56, 56, 60, 56, 57, 56, 51,...
## $ Min_Dewpoint_F            <int> 56, 56, 54, 53, 58, 51, 56, 54, 45,...
## $ Max_Humidity              <int> 93, 90, 93, 87, 93, 84, 90, 90, 86,...
## $ Mean_Humidity             <int> 75, 70, 75, 68, 77, 65, 72, 74, 58,...
## $ Min_Humidity              <int> 57, 50, 57, 49, 61, 46, 53, 57, 29,...
## $ Max_Sea_Level_Pressure_In <dbl> 30.07, 30.05, 30.00, 29.96, 29.97, ...
## $ Mean_Sea_Level_Pressure_In <dbl> 30.02, 30.00, 29.96, 29.93, 29.94, ...
## $ Min_Sea_Level_Pressure_In <dbl> 29.97, 29.93, 29.92, 29.91, 29.90, ...
## $ Max_Visibility_Miles      <int> 10, 10, 10, 10, 10, 10, 10, 10, 10,...
## $ Mean_Visibility_Miles     <int> 10, 10, 10, 10, 10, 10, 10, 10, 10,...
## $ Min_Visibility_Miles      <int> 10, 7, 10, 10, 6, 10, 10, 10, 10, 1...
## $ Max_Wind_Speed_MPH        <int> 23, 29, 26, 25, 23, 24, 29, 31, 24,...
## $ Mean_Wind_Speed_MPH       <int> 11, 13, 15, 13, 12, 15, 19, 21, 8, ...
## $ Max_Gust_Speed_MPH        <int> 28, 35, 31, 29, 30, 31, 35, 37, 28,...
## $ Precipitation_In          <fctr> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0...
## $ Cloud_Cover               <int> 4, 2, 4, 4, 6, 2, 4, 3, 0, 1, 2, 1,...
## $ Events                    <fctr> , , , , , , , , , , , , , Fog, ,...
## $ Wind_Dir_Degrees          <int> 286, 291, 284, 284, 277, 276, 269, ...
## $ zip                       <int> 94107, 94107, 94107, 94107, 94107, ...


## Observations: 70
## Variables: 7
## $ station_id   <int> 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 2...
## $ name         <fctr> San Jose Diridon Caltrain Station, San Jose Civi...
## $ lat          <dbl> 37.32973, 37.33070, 37.33399, 37.33141, 37.33672,...
## $ long         <dbl> -121.9018, -121.8890, -121.8949, -121.8932, -121....
## $ dockcount    <int> 27, 15, 11, 19, 15, 15, 15, 15, 15, 19, 19, 15, 1...
## $ landmark     <fctr> San Jose, San Jose, San Jose, San Jose, San Jose...
## $ installation <fctr> 8/6/2013, 8/5/2013, 8/6/2013, 8/5/2013, 8/7/2013...


## Observations: 18,342,210
## Variables: 4
## $ station_id      <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2...
```

```
## $ bikes_available <int> 12, 12, 12, 12, 12, 12, 12, 12, 13, 13, 13, 13...
## $ docks_available <int> 15, 15, 15, 15, 15, 15, 15, 15, 14, 14, 14, 14...
## $ time            <fctr> 2014-03-01 00:00:02, 2014-03-01 00:01:03, 201...

## Observations: 171,792
## Variables: 11
## $ Trip.ID         <int> 432946, 432945, 432944, 432942, 432941, 432940...
## $ Duration        <int> 406, 468, 534, 1041, 1091, 1126, 1135, 1151, 6...
## $ Start.Date      <fctr> 8/31/2014 22:31, 8/31/2014 22:07, 8/31/2014 2...
## $ Start.Station   <fctr> Mountain View Caltrain Station, Beale at Mark...
## $ Start.Terminal  <int> 28, 56, 56, 60, 60, 60, 60, 60, 60, 60, 61, 61...
## $ End.Date        <fctr> 8/31/2014 22:38, 8/31/2014 22:15, 8/31/2014 2...
## $ End.Station     <fctr> Castro Street and El Camino Real, Market at 4...
## $ End.Terminal    <int> 32, 76, 76, 74, 74, 74, 74, 74, 50, 50, 66, 66...
## $ Bike..          <int> 17, 509, 342, 603, 598, 555, 425, 406, 542, 62...
## $ Subscriber.Type <fctr> Subscriber, Customer, Customer, Customer, Cus...
## $ Zip.Code        <fctr> 94040, 11231, 11231, 94521, 94521, 94521, 945...

## Observations: 920
## Variables: 24
## $ PDT                      <fctr> 3/1/2014, 3/2/2014, 3/3/2014, 3/4/2...
## $ Max.TemperatureF         <int> 69, 61, 61, 66, 66, 65, 66, 72, 71, ...
## $ Mean.TemperatureF        <int> 62, 57, 56, 61, 60, 58, 57, 59, 63, ...
## $ Min.TemperatureF         <int> 54, 53, 50, 55, 54, 51, 48, 46, 55, ...
## $ Max.Dew.PointF           <int> 54, 55, 52, 54, 58, 57, 50, 51, 54, ...
## $ MeanDew.PointF           <int> 48, 51, 49, 52, 54, 53, 47, 48, 51, ...
## $ Min.DewpointF            <int> 44, 47, 47, 49, 48, 48, 43, 39, 48, ...
## $ Max.Humidity             <int> 78, 83, 89, 86, 93, 93, 89, 93, 80, ...
## $ Mean.Humidity            <int> 63, 75, 77, 77, 73, 80, 67, 63, 69, ...
## $ Min.Humidity             <int> 48, 67, 64, 67, 52, 67, 44, 33, 57, ...
## $ Max.Sea.Level.PressureIn <dbl> 29.82, 30.17, 30.15, 30.12, 30.11, 3...
## $ Mean.Sea.Level.PressureIn <dbl> 29.63, 30.01, 30.11, 30.09, 30.06, 3...
## $ Min.Sea.Level.PressureIn <dbl> 29.47, 29.83, 30.04, 30.06, 29.99, 3...
## $ Max.VisibilityMiles      <int> 10, 10, 10, 10, 10, 10, 10, 10, 10, ...
## $ Mean.VisibilityMiles     <int> 10, 9, 10, 10, 8, 9, 10, 10, 10, 10,...
## $ Min.VisibilityMiles      <int> 10, 2, 9, 7, 2, 4, 9, 10, 10, 9, 10,...
## $ Max.Wind.SpeedMPH        <int> 17, 20, 18, 17, 20, 15, 14, 17, 13, ...
## $ Mean.Wind.SpeedMPH       <int> 10, 9, 6, 5, 6, 8, 6, 4, 6, 8, 11, 1...
## $ Max.Gust.SpeedMPH        <int> 25, 26, 28, NA, 23, 20, 24, 48, 18, ...
## $ PrecipitationIn          <fctr> 0.03, 0.02, 0.1, T, 0.26, 0.01, 0, ...
## $ CloudCover               <int> 7, 6, 7, 7, 8, 6, 2, 5, 7, 5, 1, 0, ...
## $ Events                   <fctr> Rain, Rain, Rain, Rain, Rain, Rain,...
## $ WindDirDegrees           <int> 108, 181, 135, 16, 211, 258, 296, 28...
## $ Zip                      <int> 94107, 94107, 94107, 94107, 94107, 9...

## Observations: 70
## Variables: 7
## $ station_id   <int> 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 2...
## $ name         <fctr> San Jose Diridon Caltrain Station, San Jose Civi...
## $ lat          <dbl> 37.32973, 37.33070, 37.33399, 37.33141, 37.33672,...
## $ long         <dbl> -121.9018, -121.8890, -121.8949, -121.8932, -121....
## $ dockcount    <int> 27, 15, 11, 19, 15, 15, 15, 15, 15, 19, 19, 15, 1...
## $ landmark     <fctr> San Jose, San Jose, San Jose, San Jose, San Jose...
## $ installation <fctr> 8/6/2013, 8/5/2013, 8/6/2013, 8/5/2013, 8/7/2013...
```

```
## Observations: 36,647,622
## Variables: 4
## $ station_id      <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2...
## $ bikes_available <int> 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, 15...
## $ docks_available <int> 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12...
## $ time            <fctr> 2014-09-01 00:00:03, 2014-09-01 00:01:02, 201...


## Observations: 354,152
## Variables: 11
## $ Trip.ID         <int> 913460, 913459, 913455, 913454, 913453, 913452...
## $ Duration        <int> 765, 1036, 307, 409, 789, 293, 896, 255, 126, ...
## $ Start.Date      <fctr> 8/31/2015 23:26, 8/31/2015 23:11, 8/31/2015 2...
## $ Start.Station   <fctr> Harry Bridges Plaza (Ferry Building), San Ant...
## $ Start.Terminal  <int> 50, 31, 47, 10, 51, 68, 51, 60, 56, 47, 60, 67...
## $ End.Date        <fctr> 8/31/2015 23:39, 8/31/2015 23:28, 8/31/2015 2...
## $ End.Station     <fctr> San Francisco Caltrain (Townsend at 4th), Mou...
## $ End.Terminal    <int> 70, 27, 64, 8, 60, 70, 60, 74, 55, 66, 77, 70,...
## $ Bike..          <int> 288, 35, 468, 68, 487, 538, 363, 470, 439, 472...
## $ Subscriber.Type <fctr> Subscriber, Subscriber, Subscriber, Subscribe...
## $ Zip.Code        <fctr> 2139, 95032, 94107, 95113, 9069, 94118, 92562...


## Observations: 1,825
## Variables: 24
## $ PDT                     <fctr> 9/1/2014, 9/2/2014, 9/3/2014, 9/4/2...
## $ Max.TemperatureF        <int> 83, 72, 76, 74, 72, 72, 72, 68, 72, ...
## $ Mean.TemperatureF       <int> 70, 66, 69, 68, 66, 66, 66, 64, 65, ...
## $ Min.TemperatureF        <int> 57, 60, 61, 61, 60, 60, 60, 59, 57, ...
## $ Max.Dew.PointF          <int> 58, 58, 57, 57, 57, 55, 54, 52, 56, ...
## $ MeanDew.PointF          <int> 56, 57, 56, 57, 56, 54, 53, 51, 53, ...
## $ Min.DewpointF           <int> 52, 55, 55, 56, 54, 52, 45, 50, 52, ...
## $ Max.Humidity            <int> 86, 84, 84, 84, 84, 84, 78, 72, 87, ...
## $ Mean.Humidity           <int> 64, 73, 69, 71, 71, 69, 66, 64, 72, ...
## $ Min.Humidity            <int> 42, 61, 53, 57, 57, 53, 53, 55, 57, ...
## $ Max.Sea.Level.PressureIn  <dbl> 29.86, 29.87, 29.81, 29.81, 29.92, 2...
## $ Mean.Sea.Level.PressureIn <dbl> 29.82, 29.82, 29.76, 29.76, 29.87, 2...
## $ Min.Sea.Level.PressureIn  <dbl> 29.76, 29.79, 29.72, 29.72, 29.81, 2...
## $ Max.VisibilityMiles     <int> 10, 10, 10, 10, 10, 10, 10, 10, 10, ...
## $ Mean.VisibilityMiles    <int> 10, 10, 10, 10, 9, 9, 10, 10, 10, 9,...
## $ Min.VisibilityMiles     <int> 8, 7, 10, 8, 7, 7, 10, 10, 10, 5, 2,...
## $ Max.Wind.SpeedMPH       <int> 16, 21, 21, 22, 18, 17, 18, 18, 17, ...
## $ Mean.Wind.SpeedMPH      <int> 7, 8, 8, 8, 8, 9, 10, 12, 7, 5, 7, 8...
## $ Max.Gust.SpeedMPH       <int> 20, NA, 24, 25, 32, 30, 28, 22, 21, ...
## $ PrecipitationIn         <fctr> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ CloudCover              <int> 0, 5, 4, 5, 4, 4, 3, 4, 5, 1, 2, 1, ...
## $ Events                  <fctr> , , , , , , , , , , , , , , Rai...
## $ WindDirDegrees          <int> 290, 290, 276, 301, 309, 290, 293, 2...
## $ Zip                     <int> 94107, 94107, 94107, 94107, 94107, 9...
```