1. Data cleaning / Feature engineering
   1. Both tables
      1. Remove rows with empty values from both tables
         1. For the b-cycle one, can we add in the checkout\_kiosk\_id from the name.
      2. Truncate dates
   2. B-cycles
      1. Trim rides that last longer and shorter (from the same station) than a certain amount.
      2. Trim membership types and things like that?
      3. Possibly trim year?
      4. Possible features to delete
         1. Trip id
         2. Bicycle id
         3. Yearmonth
   3. Weather
      1. PCA? Remove some features
         1. Dewpoint
         2. Remove everything but average humidity
         3. Sea level pressure
         4. Average visibility
         5. Only keep average wind speed
      2. Precipitation: chance the trace to something
      3. Events: only like rain, fog, thunderstorm, snow
         1. Change to rain Y/ N and fog Y / N
   4. Join weather and B-cycle tables on date.
2. Data mining preparation
   1. Distances between stations
      1. Could we do this with google maps? Could we do this with latitude longitude
   2. Distance in dates could be modulo.
   3. Standardize / normalize
3. Data mining ideas
   1. Unsupervised learning: do we get any interesting clusters?
   2. Supervised learning
      1. Picking a station & the time and such, can we predict where they’re going / how long the trip will be.