**Instructional Days:** 7-10

**Topic Description:** This lesson introduces R/Deducer as a data analysis tool. The basic features of loading and saving files, sorting and creating subsets are explored. Maps are created by using the latitude and longitude of a location and then maps of points and bubble charts are created from a file of data.

**Objectives:**

The students will be able to:

* Translate addresses into latitude/longitude.
* Sort files of data.
* Create subsets of data.
* Read location data from a file and plot points on maps.
* Create bubble plots.

**Outline of the Lesson:**

* Journal Entry (5 minutes)
* Describing location (30 minutes)
* Exploring LA Bike Data and Deducer (30 minutes)
* LA Bike Activity (45 minutes)
* LA Bus Stops Activity (45 minutes)
* Bubble charts (20 minutes)
* Bubble Charts Activity (45 minutes)

**Student Activities:**

* Complete journal entry.
* Participate in discussion of location, LA Bike Data and Deducer.
* Complete LA Bike Activity.
* Complete LA Bus Stops Activity.
* Participate in discussion of bubble charts.
* Complete Bubble Chart Activity.

**Teaching/Learning Strategies:**

* Journal Entry: Consider the data that you have been collecting with the phone. How might seeing the data on a map help you analyze it?
* Install Deducer package. (See Deducer Quick Start Guide.)
  + You may want to do this installation yourself before the class to save time.
* Describing location
  + Use Walking and Biking in LA as an introduction to the LA bike data and describing location with latitude and longitude. (You may want to share a version of this resource with students.)
    - Take the opportunity to point out that this is a campaign similar to what they will be doing for the final project—people like them concerned about a topic, collecting data to inform their cause.
  + Load the Google Map with the locations of the 56 intersections for the survey.
  + Ask questions about these locations: Are any of these locations in your neighborhood? Near the school? Near your home? Other questions that may be of interest.
  + Go to <http://www.getlatlon.com/> and demonstrate how to translate a place on the map to latitude and longitude.
    - Type in an address. Use the example in Walking and Biking in LA Teacher Resource or another that you choose
    - The appropriate numbers are under the map inside the parenthesis next to WKT: POINT( - 118.2796304, 34.0916803). That means the location is 34.0916803 north of the equator and 118.2796304 west of the Prime Meridian. (If the example provided is used.) Note: Latitude is often referred to first, but the coordinates of the point are (longitude, latitude).
    - Have students try finding locations for their house and/or school.
    - Point out that the system of longitude and latitude will allow them to draw spatial objects such as points (house, school). They can think of longitude as the x-direction (it runs along the equator from east to west) and latitude as the y-direction (it runs from the North Pole to the South Pole along the Prime Meridian) when making a plot from these coordinates.
  + Exploring LA Bike data and Deducer
    - Load the labike data file.
    - Point out that this is what they will be doing with the data they collect with the phones.
    - Point out the following features as you discuss the layout of the table shown in the data viewer (You can use Exploring LACBC and Deducer as a reference.):
      * Header, number of rows, categories
      * Note what you see in the data viewer (many of rows of data).
      * The first line is the header and describes the names of each variable or column.
      * Each row refers to a different intersection, and so there are 38 intersections represented in the data set. Each column refers to the various data that were collected about the intersection.
      * Navigate through the survey to show the variables in the data set:
        + "name" is the location of the intersection,
        + "longitude" is the longitude of the location
        + "latitude" is the latitude of the location
        + "type" is the type of bike transportation available at the intersection (bike lane, bike path, bike route, none)
        + “bike\_count\_pm” is the evening count of bikes
        + “ped\_count\_pm” is the evening count of pedestrians
      * Demo how to obtain a table of frequencies for type.
        + The table appears in the Console window.
        + Note that 20 of the intersections have nothing.
        + There will be more discussion of frequencies later in the unit, but note that this file is small enough that the counting could be done by hand (as with the data collected in Unit 2), but that later data sets will be much larger.
      * Demo how to sort by bike count and pedestrian count.
        + Which intersections have the most bike traffic/ pedestrian traffic? Are they the same?
      * Demo how to create a subset of locations where the bike count is greater than or equal to the pedestrian count.
        + Point out that the system provides a default title for the subset, but it is better to create a new title.
        + Ask students what other subsets might be interesting to create (e.g., locations with bike routes).
        + Have students create a few subsets of their own and list the questions they might want to ask about those subsets.
      * Demo how to plot the intersections on a map. Include a title, axes and background. Also demo the various sizes and shapes of points and how to zoom.
        + Before each plot feature is added, ask students questions that will guide them to the need for the feature.
        + Ask students questions about the plot such as: Are there any outliers? Are there clusters of points? Does the plot match the table?
  + LA Bike Activity
    - Have students complete the LA Bike Activity on their own. Circulate the room and answer questions.
    - Allow sufficient time at the end of this part to ask students for their responses and lead a discussion to ensure that they understand each of the features of Deducer discussed so far.
  + LA Bus Stops Activity
    - Have students complete the LA Bus Stops Activity on their own. Circulate the room and answer questions.
    - Allow sufficient time at the end of this part to ask students for their responses and lead a discussion to ensure that they understand each of the features of Deducer discussed so far.
  + Bubble charts with LA Bike Count Data
    - Describe bubble charts. (You can use Bubble Charts as a resource.)
    - Ask questions such as: What is being described when longitude and latitude is plotted on a map? Is there a way to distinguish counts of pedestrians and bicyclists?
    - Demo how to create a bubble chart with the pedestrian counts in the labike file.
    - Demo how to change the size and color of the bubbles.
  + Bubble Chart Activity
    - Have students complete the Bubble Charts Activity on their own. Circulate the room and answer questions.
    - Allow sufficient time at the end of this part to ask students for their responses and lead a discussion to ensure that they understand each of the features of Deducer discussed so far.
* Deducer Quick Start Guide
* Walking and Biking in LA
* Exploring LACBC and Deducer
* LA Bike Activity
* LA Bus Stops Activity
* Bubble Charts
* Bubble Charts Activity

**Walking and Biking in LA**

**\*\*Survey Description**

The data that will be considered first in this lesson were collected in September of 2009 by the Los Angeles County Bicycle Coalition (LACBC, http://la-bike.org), a non-profit organization that works to "make the entire L.A. region a safe and enjoyable place to ride." For two days in late September, the LACBC recruited volunteers to count the number of bicyclists and pedestrians that pass 56 different intersections within Los Angeles County. Some of the survey locations were chosen because they are known to be popular with cyclists and pedestrians, others because they are near locations where a traffic-related change is about to take place, and still others because they are the site of a large number of bike accidents each year.

LACBC volunteers surveyed each location in the morning (7:00-9:30 am) and evening (4:00-6:30 pm) on Tuesday September 22 and Wednesday September 23 of 2009. Data were also collected on Saturday the 26th, but will not be considered here. The volunteers produced a report summarizing their findings (<http://lacbc.files.wordpress.com/2010/06/labikecountreport.pdf>).

**\*\* Describing location**

When someone asks you to describe your current location, you might respond informally by saying you are in class or at school. Friends and family will know where that is, but if a relative were visiting from out of town and were unfamiliar with the area, the street address for school and maybe a nearby intersection would be necessary. These descriptions are excellent for looking up a location on a map or for walking, biking, or driving somewhere new. Roads and intersections and street addresses create a network that we regularly navigate. This network may change as old roads and buildings are replaced by others. Also, some of the important places in our lives do not have a street address (like the peak of a mountain or a hiking trail in the Santa Monica Mountains). Finally, in order to draw a map of LA, it would be helpful to be able to specify positions in a more consistent way, tracking a road as it turns a corner or veers to the left. For all of these reasons, there is a need to associate positions with a fixed set of "coordinates" on the earth.

One of the most popular such coordinate systems involves specifying a point's latitude and longitude. These are two numbers that represent angles (in degrees) from the center of the earth to a point on its surface. Latitudes are angles from north to south—in this case the North Pole is assigned a value of 180 degrees, the equator is at 0 degrees and the South Pole is at -180 degrees. Longitudes are angles from east to west with 0 occurring at the Prime Meridian, a line running from the North Pole to the South Pole and crossing Greenwich, England. (Greenwich is also used in defining Greenwich Mean Time or GMT.) A description of longitude and latitude can be found at the following url.

<http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=Georeferencing_and_coordinate_systems>

These two ways of describing a location (an address versus latitude-longitude) can be compared. The LACBC produced a Google Map with the locations of the 56 intersections for their survey—this was actually a recruitment tool and there are references to shifts that are needed for people to lend a hand.

<http://tinyurl.com/LABikeECS>

To translate the location on the map for the intersection of Sunset and Hyperion into latitude and longitude, a service like Get Lat Lon (<http://www.getlatlon.com/>) can be used.

In the box at the top of the page enter

Sunset & Hyperion, Los Angeles, CA

The map should center on the right intersection. The latitude and longitude of the point will appear at the bottom of the page. In this case the latitude and longitude of the point is 34.0916803, -118.2796304—that is, 34.0916803 north of the equator and 118.2796304 west of the Prime Meridian.

The system of longitude and latitude allows people to draw spatial objects like points (house, school) or lines (a street) or shapes (the grounds of a high school, a park). Longitude can be thought of as the x-direction (it runs along the equator from east to west) and latitude as the y-direction (it runs from the North Pole to the South Pole along the Prime Meridian) when making a plot from these coordinates.

Load the labike data file into the Deducer Data Viewer.

It should appear similar to the table below.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| name | longitude | latitude | type | bike\_count\_pm | ped\_count\_pm |
| 1st & Alameda | -118.238125 | 34.049175 | none | 62 | 241 |
| 4th & Wilton | -118.313441 | 34.06713 | bike route | 48 | 87 |
| 7th & Figueroa | -118.259883 | 34.049388 | none | 216 | 1979 |
| 8th & La Brea | -118.344641 | 34.060446 | none | 72 | 272 |
| 9th & Pacific | -118.287306 | 33.735118 | none | 58 | 160 |

The first line of the file is known as a header and describes the names of each variable or column (e.g., "name", "longitude", and "latitude"). Each row refers to a different location at which volunteers counted the bike and pedestrian counts during the evening rush hour. ("Objects" in this table are positions in Los Angeles and the "variables" measured for each include the name, the longitude and latitude, and the counts of pedestrians and cyclists.)

The data frame "labike" has 38 rows, each referring to a different location. The first column is the name of the locations, similar to the list in the LACBC Google Map. (The data here reproduce Table 14 from LACBC's report and there are only 38 of the 56 locations included.) The next two columns give the positions' longitude and latitude. These coordinates can be used to place the locations on a map. The fourth column describes the type of bike transportation available at the intersection (a bike lane, a bike path, a bike route or nothing) and the last two columns represent the evening counts of bikes (column 5) and pedestrians (column 6) crossing the intersection.

A variety of operations can be performed on the data. (e.g., create a table that shows the frequency of type, sort by bike\_count\_pm , sort by ped\_count\_pm, create a subset of locations where the bike count is greater than the pedestrian count ( bike\_count\_pm ≥ ped\_count\_pm)).

There are several kinds of spatial data. Their structure is best described by their look—that is, positions or points on the map (a house, the Sandstone Peak in the Santa Monica Mountains); points that are connected to form paths or lines (the route of a walk to school or the driving route to a friend's house); and areas or regions (the footprint of a school's buildings or the area covered by Los Angeles County). Points, lines and regions are basic spatial structures that are used for computation.

In the case of the LA Bike Count data, there are intersections where survey takers stood (points). The transportation system in Los Angeles can also be consulted for bus routes (lines), and the U.S. Census Bureau can provide demographic data about people living in different Census blocks (small geographic areas or regions).

Since the data set includes the longitude and latitude in columns 2 and 3, the intersections can be plotted on a map.

Load the labike data file into the Deducer Data Viewer.

1. Create a subset of the locations with no special routes for bikes.

* How many locations are in the subset?
* Sort by bike\_count\_pm. Which intersection has the greatest count? Which has the least count?
* Plot this subset on a map. Include a title, axes and a background.

1. Create a subset of the locations with special routes for bikes.

* Use a different color and shape to plot this subset on the same map.
* Describe any patterns you see.

Load the bus\_stops data file into the Deducer Data Viewer.

1. What are the variables in this survey?
2. Form a frequency table to see the number of stops along each street. Which street has the most stops? What might be a reason for this?
3. How many total stops are there?
4. Look at the data for the 6000th row—a bus stop on Sunset Boulevard at Anita Avenue. Go to http://getlatlon.com and type in Sunset & Anita, Los Angeles, CA and check that the longitude and latitude listed in the data file are the same as from Get Lat Lon. What do you notice? Why might this be the case?
5. Create a subset of the bus stops that are along Sunset or Vermont. How many stops are there?
6. Create a plot of bus stops that are along Sunset or Vermont. Include a title, axes, and a background. Describe what you see in the plot.
7. Create a plot of bus stops that are along Myrtle or Mulholland. Include a title, axes, and a background. Describe what you see in the plot. How does this compare to the plot of Sunset or Vermont.
8. Create a plot of bus stops that are along Gayley or Hilgard. Include a title, axes, and a background. Describe what you see in the plot. How does this compare to the previous plots?
9. Create a plot of bus stops that are along a few streets in your neighborhood. Include a title, axes, and a background. Describe what you see in the plot. Why might this data be useful for someone to have?
10. What is the advantage to plotting the data on a map instead of just looking at the latitude/ longitude numbers?
11. What would happen to the map if you had less data in the file? More data in the file? How would you affect your interpretation of the map?
12. If you were trying to make a case that you needed more bus stops in your neighborhood would it be enough to show that the count of bus stops is less than those along Sunset? Explain your answer.
13. How could you use what you learned about plotting points on a map with the data collected on the phone?

**Bubble Charts**

The points at which LA Bike volunteers stood and the position of bus stops exhibit the geometry of these things. However, spatial objects can have other data associated with them. The LA Bike Counts are associated with counts of pedestrians and bicyclists. When the intersections at which the LA Bike volunteers stood were plotted there was nothing that could be determined about the number of bikers or pedestrians.

A bubble chart uses numerical values to scale the diameter of circles located at a given spatial location. Consider the pedestrian totals from the LA Bike data.

In a bubble plot of the pedestrian counts each intersection where volunteers collected data is the center of a circle—the larger the circle, the greater the number of pedestrians counted there. If this plot were drawn by hand, a number of choices would need to be made. First, the size of the circles relative to each other is fixed by the data. If a volunteer at one intersection saw twice as many pedestrians as another volunteer saw at a different intersection, the first circle should be twice as big as the second. The relationship between the circles and the map, over which they are plotted, however, is not fixed and can be changed (again, assuming the relative sizes of the circles remains the same).

**Bubble Chart Activity**

Load the labike data file into the Deducer Data Viewer.

1. Create a bubble chart of the pedestrian counts.
2. Create a bubble chart of bike counts and add it to the pedestrian counts. What happens to the chart?
3. Change the color for bike counts. Describe what you see now.
4. What else might you change to get an even clearer visual picture of bike and pedestrian counts? Try these ideas. Explain how this changes the chart.
5. Based on your graph, what questions might you ask?
6. Try zooming in on a part of the graph. Describe what you see.
7. Does your graph make sense based on the counts in the table? Explain why or why not.
8. Create a subset of all locations that have a special route for bikes. Create a bubble chart with the counts for the subset. Describe what you see.
9. Create a subset of all locations that have no special route for bikes. Add a bubble chart of the counts for the subset to the previous chart in a different color. Describe what you see now. What conclusions might you draw? Justify your answer.
10. Create another pair of subsets that are of interest to you. Create a bubble chart. Describe the story you see.