James Li

11.007

Prof. David Hsu

December 14th, 2018

Lab Report 3

Introduction

The purpose of this lab report is to first use R (with Shiny and Leaflet) to visualize data gathered during our class bike ride on December 6th, and then to use this visualization to examine the dataset. Temperature, humidity, particulate matter, and location data was gathered for a period of about 45 minutes by multiple sensors mounted onto bikes, which the class rode along different routes in the Cambridge area. A number of GPS sensors proved to be unreliable and had trouble finding satellites, so only a subset of the total sensors that were deployed are examined in this report.

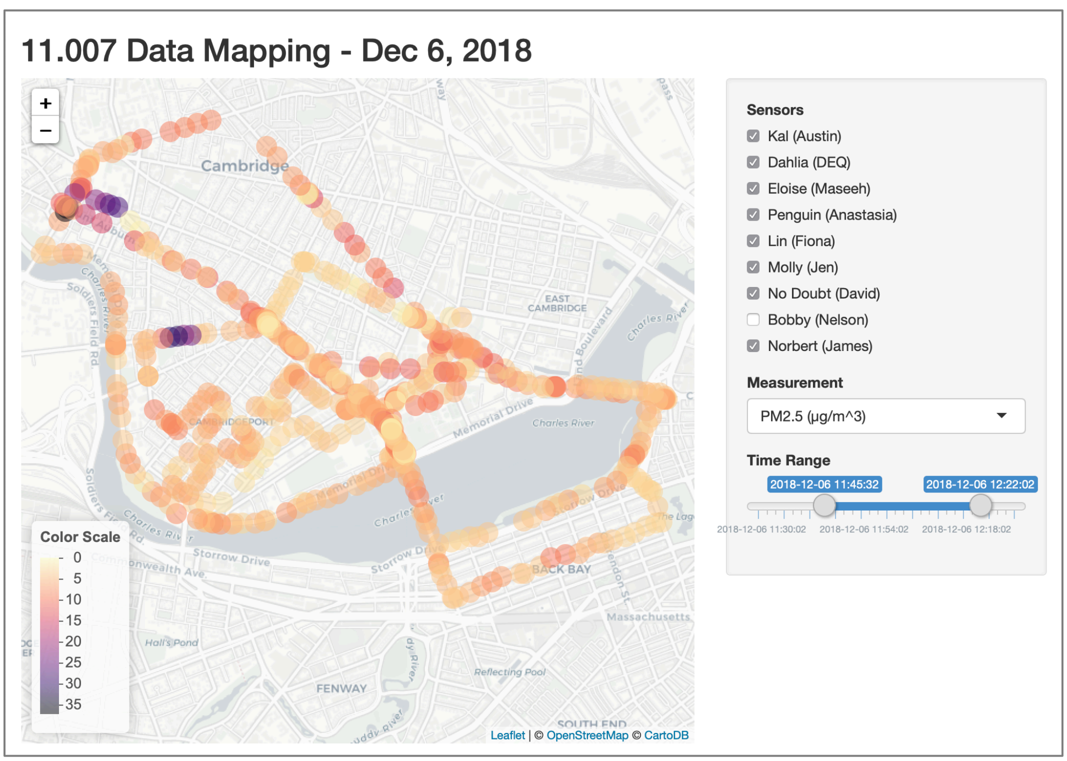

Figure 1: Screenshot of Shiny data visualization app. Note the filtering controls at right and color scale legend at bottom left.

Visualizing with Shiny and Leaflet

My Shiny app allows for the data to be filtered by sensor, measurement (e.g. temperature, humidity, PM1…), and time range. Data from multiple sensors can be viewed simultaneously, but for simplicity only one measurement can be viewed at once. A color scale is used to show measurement values on the map, with lighter colors corresponding to lower values and darker colors corresponding to higher values. I chose to keep the radii of the points on the map constant, in order to avoid confusion about the physical area represented by each data point. A screenshot of the app is shown in Figure 1.

Temperature and Relative Humidity

Examining the temperature data mainly shows trends in time, but also suggests a few differences between different locations. The sensors begin indoors in the class lab space, where the temperature is around 78ºF; when they leave the lab space, they begin to measure much lower temperatures. The air temperature for Cambridge at the time of the bike ride was around 38ºF (darksky.net); the sensor Eloise (stationary at Maseeh Hall) measured a fairly constant 41-43ºF during this time. It takes several minutes for the measured temperatures to fall into the low 40s; this is visible on the map as a decrease in temperature along the paths of the sensors (see Figure 2). This trend is most likely due to the slow thermal response of the sensors (as the enclosure and internal parts cool over time and approach the outside temperature), and not due to actual temperature differences in space. As shown in Figure 2, NoDoubt passes the intersection of Vassar Street and Massachusetts Avenue less than eight minutes after Penguin, but reports a temperature 10ºF lower, supporting this explanation. By the time all of the sensors return to MIT, they are reporting temperatures within about 2ºF of Eloise.

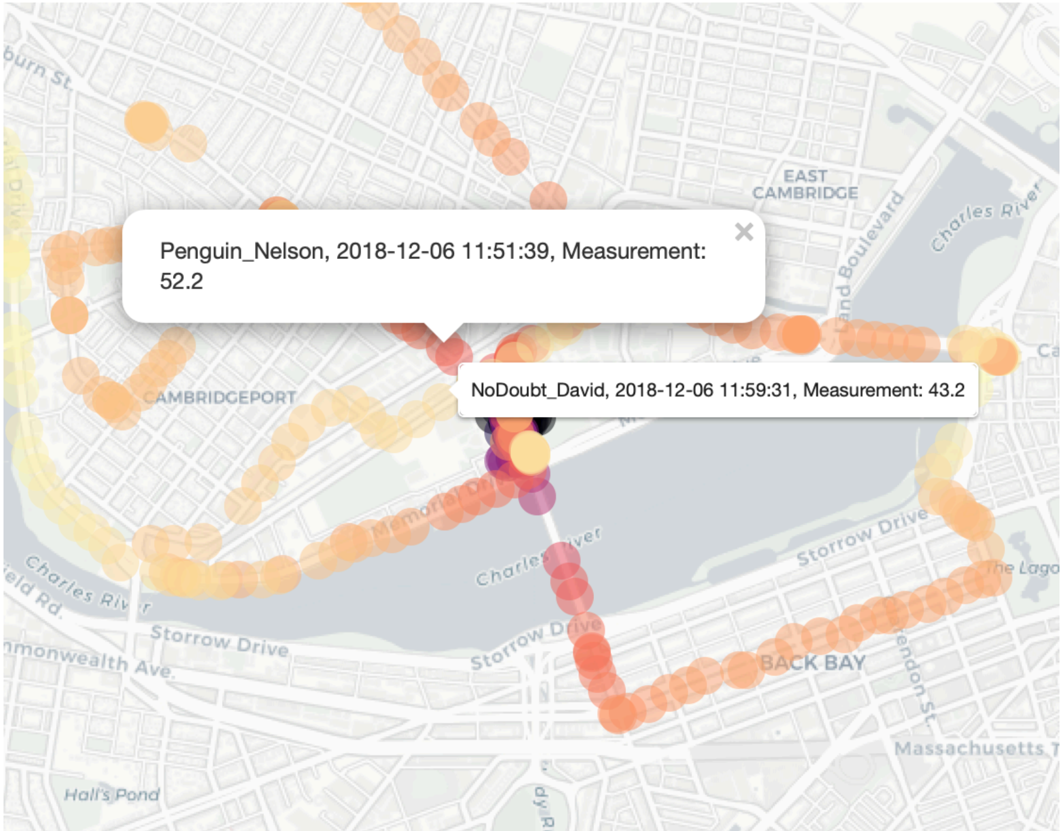

Figure 2: Disparity in temperature measurements at the same point in space but different times, likely due to slow thermal response of sensors. Note also the general trend for measured temperatures to decrease as the sensors move away from MIT.

However, beyond this trend in time, the data shows several differences between sensors. For example, at around noon, sensors Molly and Penguin pass within 750 feet of each other, but Penguin reports a temperature that is about 4ºF higher than that reported by Molly. This is shown in Figure 3. When Molly meets Norbert a few minutes later, the two sensors' temperature readings are within 0.5ºF of each other, suggesting that Molly is not simply reporting unusually low temperatures. This data may point at the possibility of lower temperatures along the Charles River and a slight heat island effect in the Cambridgeport area, but it is also possible that Penguin simply had a slower thermal response and took more time to cool down from indoor temperature (e.g. due to internal design or mounting configuration).

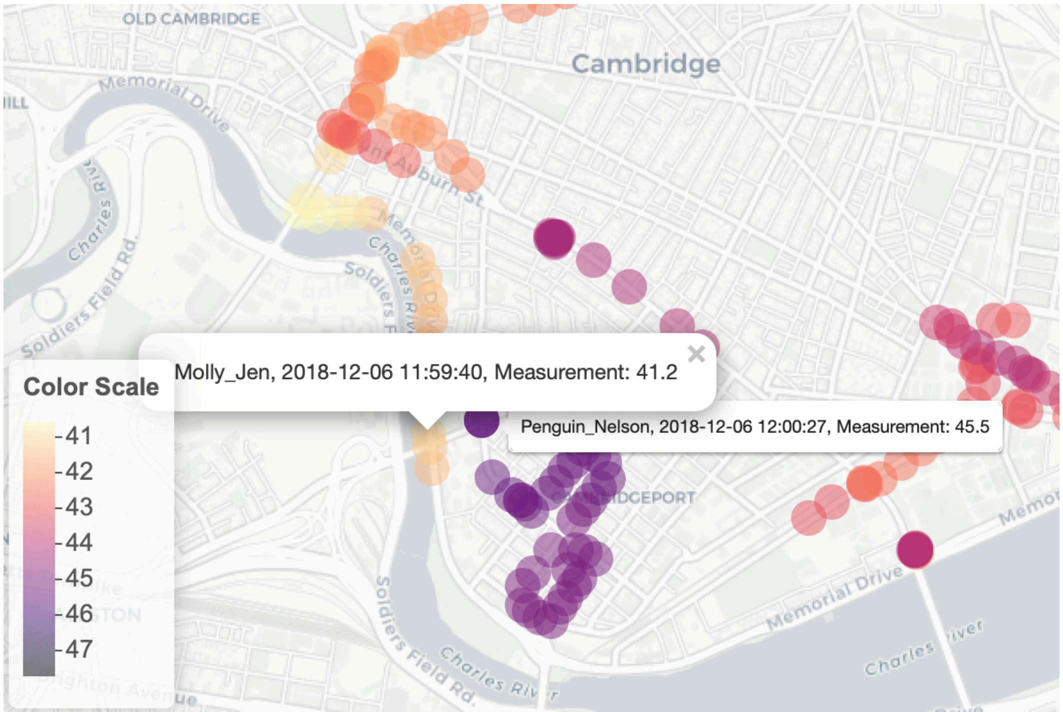

Figure 3: Temperature disparity between Molly and Penguin. Data points displayed are during the time period 11:59:02 to 12:09:02. Note the generally cooler temperatures recorded by Molly, and warmer temperatures reported by Penguin.

Similarly, during the data collection period the stationary sensor DahliaDEQ in Roxbury measured fairly consistent temperatures around 46ºF, about three degrees higher than the temperatures measured by Eloise (stationary at Maseeh Hall) and the other sensors in Cambridge. This may also support the hypothesis that there is a heat island effect away from bodies of water; however, it is also possible that other factors affected the temperature reading (e.g. exposure to direct sunlight). Without more data in this area, it is difficult to draw meaningful conclusions.

The relative humidity data is generally inversely related to the temperature data. Where Penguin recorded higher temperatures as described above, the same sensor reported lower relative humidities (4-8% lower) compared to other sensors. One interesting disparity is in the data for the stationary sensor Eloise, which reported a consistently much higher (5-15% higher) humidity compared to both the moving sensors and the other stationary sensor (Dahlia) in Roxbury. Figure 4 shows this data.

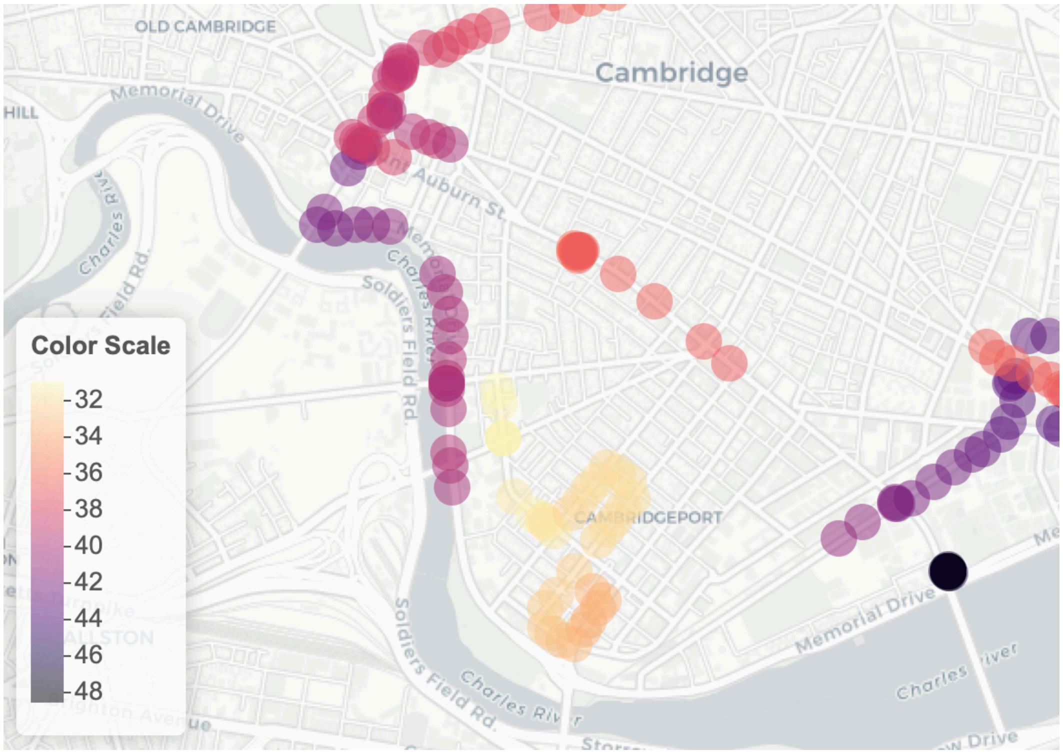

Figure 4: Relative humidity data for the same time period as in Figure 3. Note the lower humidities reported by Penguin (in Cambridgeport) and the higher humidity reported by Eloise (at Massachusetts Avenue and Memorial Drive).

Particulate Matter

For this section, I will focus on examining the PM2.5 data, as the three particulate matter readings (PM1, PM2.5, PM10) are generally very similar. On the whole, particulate matter readings across Cambridge were very low, remaining generally below 10-15 µg/m^3. This was reflected by the data from Eloise (Maseeh), where PM2.5 ranged from 1-6 µg/m^3, and Dahlia (Roxbury), where PM2.5 ranged from 1-7 µg/m^3. There were three locations where particulate matter readings were notably higher: construction near 950 Massachusetts Avenue, Harvard Square, and Western Avenue at Howard Street. Overall PM2.5 data is shown in Figure 5.

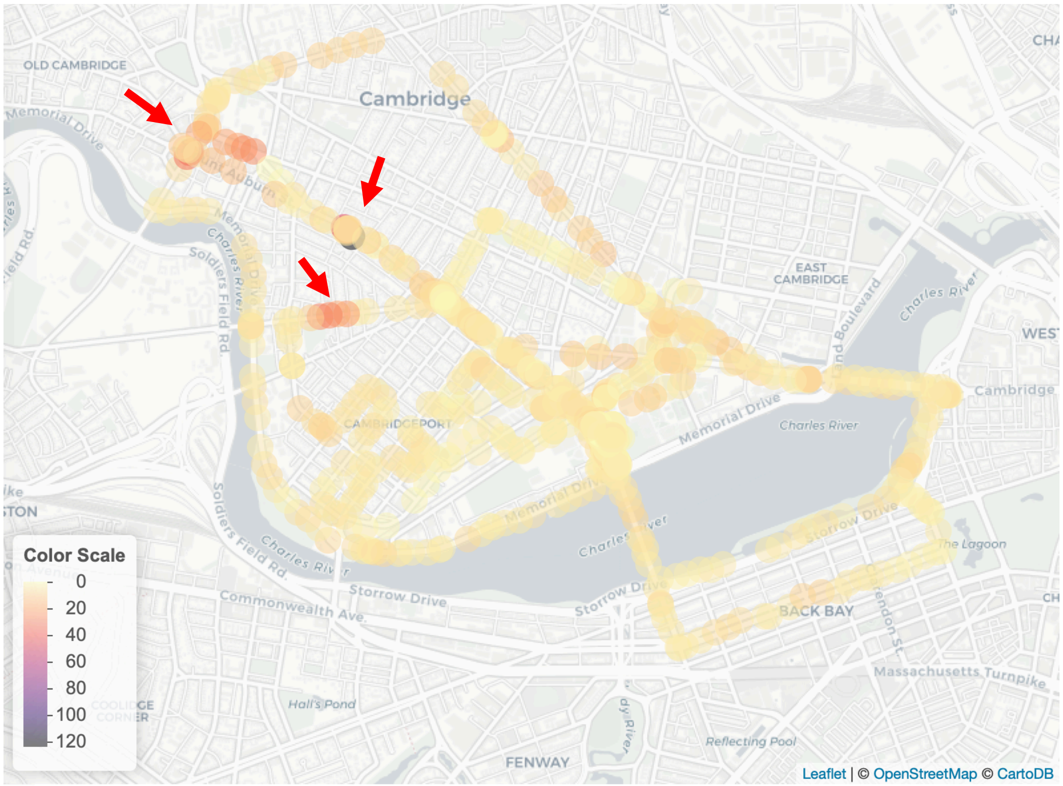

Figure 5: PM2.5 data from all sensors for the time period 11:30 to 12:30. The three locations of elevated readings are indicated with arrows.

At 950 Massachusetts Avenue, sensor Bobby was stopped for several minutes near an active construction site where asphalt was being poured and a diesel generator was running. This resulted in a peak PM2.5 reading of 135 μg/m^3, as well as several readings above 50 μg/m^3. This result is not surprising and fairly believable given the circumstances. However, sensors Norbert and Molly passing by 13 minutes later only recorded PM2.5 readings of around 10 μg/m^3. I speculate that this is due to the high speed at which Norbert and Molly were traveling; they likely would have spent only a few seconds at most passing through the high-PM area.

Harvard Square was a second site of elevated PM2.5 readings. Molly reported high values at the intersection of John F. Kennedy Street and Mt. Auburn Street (34-42 μg/m^3), as well as on Massachusetts Avenue between Bow and Holyoke Streets (25-36 μg/m^3). The values reported by Norbert were not as high but were still elevated: 16-22 μg/m^3 at various points throughout the area. Both sensors passed through Harvard Square within the same five minute window. This data is shown in Figure 6. One possible explanation for these high PM readings is that Harvard Square has more vehicle traffic than other areas, but this cannot be conclusively determined from the data.

Figure 6: Elevated PM2.5 measurements in Harvard Square, during the time period between 12:05:02 and 12:10:02. Norbert entered Harvard Square from the north and exited eastbound on Mt. Auburn Street, while Molly entered from the south on John F. Kennedy Street and exited eastbound on Massachusetts Avenue.

At Western Avenue and Harold Street, sensor Penguin reported four data points in the range 22-32 $\mu g/m^3$, increased from 5-6 $\mu g/m^3$ at other parts of Western Avenue. The multiple readings suggest that there was some source of particulate matter in that vicinity, although no other sensors passed through the same area to corroborate these readings. No further conclusions can be drawn from this data.

Future Work

This data collection provided interesting initial insights into temperature and particulate matter distribution in the Cambridge area, but more work is needed to confirm these initial results and investigate possible sources. For example, longer periods of data collection would reduce or eliminate the effects of slow thermal response in the sensors and could also reveal trends depending on time of day. A greater number of sensors could be used to create a more dense network, so that data from multiple sensors is available to corroborate or disprove local trends. Similarly, slower-moving sensors would be less affected by speed and would provide more accurate readings of each area, rather than averaged readings over a large area.

R Script

```
# Mapping class bike ride data for Lab Report 3 using Leaflet and Shiny
# Based on example scripts from class


# Clear workspace, load libraries, clean data
# ============================================

# clear variables, workspaces
rm(list=ls())
cat('\014')

# libraries
library(shiny)
library(leaflet)
library(RColorBrewer)
library(viridis)

# folder locations
setwd("~/Dropbox (MIT)/11.007/lab report 3")
path <- "~/Dropbox (MIT)/11.007/lab report 3/data downloaded 181206/"

# read in cleaned data
dat <- read.csv(paste0(path,"181206 cleaned dat.csv"))

# remove values where lat or long == 0 or pm < 0
dat <- dat [ dat$lat !=0 & dat$long != 0 & dat$pm1 >= 0 & dat$pm25 >= 0 & dat$pm10 >=
0, ]

# remove unneeded columns
dat <- subset(dat, select = -c(id, X, feed_id))

# reorder in time and add num_id column
dat <- dat [ order (dat$created_at),]
dat$num_id <- seq(1, dim(dat)[1],1)

# fix things
dat$created_at <- as.character(dat$created_at)
dat$created_at <- as.POSIXct( dat$created_at, tz = "UTC")
attributes(dat$created_at)$tzone <- "EST" # time zone shift

# truncate time range
dat <- dat[dat$created_at >= as.POSIXct("2018-12-06 11:30:00"),] # filter start time
dat <- dat[dat$created_at <= as.POSIXct("2018-12-06 12:30:00"),] # filter start time


# Support lists for Leaflet and Shiny
# ===================================
```

```
sensorNames  =  c("Kal_Nelson",  "DahliaDEQ_Nelson",  "Eloise_Jen",  "Penguin_Nelson",
"Lin_Fiona",
                "Molly_Jen",           "NoDoubt_David",           "BobbyBike_Nelson",
"NorbertBike_James")
sensorDisplayNames = c("Kal  (Austin)", "Dahlia  (DEQ)", "Eloise  (Maseeh)", "Penguin
(Anastasia)", "Lin (Fiona)",
                       "Molly(Jen)",   "NoDoubt   (David)",   "BobbyBike   (Nelson)",
"NorbertBike (James)")
colNames = c("temp", "rh", "pm1", "pm25", "pm10")


# Wrap in Shiny server
# ====================

ui <- fluidPage(
  titlePanel("11.007 Data Mapping - Dec 6, 2018"),
  sidebarLayout(
    position = "right",
    sidebarPanel(
      checkboxGroupInput("sensor", label = "Sensors",
        choices = list("Kal (Austin)" = 1,
                       "Dahlia (DEQ)" = 2,
                       "Eloise (Maseeh)" = 3,
                       "Penguin (Anastasia)" = 4,
                       "Lin (Fiona)" = 5,
                       "Molly (Jen)" = 6,
                       "No Doubt (David)" = 7,
                       "Bobby (Nelson)" = 8,
                       "Norbert (James)" = 9),
        selected = c(1,2,3,4,5,6,7,8,9)
        ),
      selectInput("measurement", label = "Measurement",
        choices = list("Temperature (ºF)" = 1,
                                 "Relative Humidity (%)" = 2,
                                 "PM1 (µg/m^3)" = 3,
                                 "PM2.5 (µg/m^3)" = 4,
                                 "PM10 (µg/m^3)" = 5),
        selected = 5),
      sliderInput("timeRange", label = "Time Range", min = min(dat$created_at),
                  max   =   max(dat$created_at),   value   =   c(min(dat$created_at),
max(dat$created_at)))
      ),
    mainPanel(
      tags$style(type = "text/css", "#map {height: calc(100vh - 80px) !important;}"),
# set height
      leafletOutput("map")
      )
  )
)

server <- function(input, output, session) {

  filtered_data <- reactive({
```

```
    # filter sensors
    smalldat <- dat[dat$sensor_id %in% sensorNames[as.numeric(input$sensor)],]

    # filter time range
    times  <-  as.POSIXct(as.character(input$timeRange),  tz="UTC")  #  convert  into
correct format
    attributes(times)$tzone <- "EST" # time zone shift
    smalldat <- smalldat[smalldat$created_at >= times[1],] # filter start time
    smalldat <- smalldat[smalldat$created_at <= times[2],] # filter end time

    # create column for selected measurement parameter
    smalldat$selected <-  smalldat[,as.numeric(input$measurement)]

    # output
    smalldat
  })

  # Create the map
  output$map <- renderLeaflet({
    leaflet() %>%
      addProviderTiles(providers$CartoDB.Positron) %>%
      setView(lng = mean (dat$long),
              lat = mean (dat$lat),
              zoom = 13)
  })

  observe({

    # create color palette, scaling based on selected data
    pal     <-     colorNumeric(palette     =     rev(magma(150)),     domain     =
range(filtered_data()$selected))

    # update map via leafletProxy
    leafletProxy("map", data = filtered_data()) %>%
      clearMarkerClusters() %>%
      clearMarkers %>% # get rid of existing markers before redrawing!
      clearControls %>%
      addCircleMarkers(~long, ~lat,
                       radius = 10,
                       color = ~pal(selected),
                       stroke = FALSE,
                       fillOpacity = 0.5,
                       popup = ~paste0(sensor_id,
                                       ", ", as.character(created_at),", Measurement:
", selected),
                       label = ~paste0(sensor_id,
                                       ", ", as.character(created_at),", Measurement:
", selected)) %>%
      addLegend("bottomleft", pal = pal, values = ~selected, title = "Color Scale")
  })
}

shinyApp(ui, server)
```