# Lab 7

Start Assignment

**Due** Monday by 11:59pm **Points** 30 **Submitting** a text entry box or a file upload

**EE5373: Data Modeling Using R**

**Summer, 2023**

Department of Electrical and Computer Engineering

**University of Minnesota**

Lab 7: Plotting the house price predictions on a map.

-

Due date: See the due date shown on the class web page.

Goal: This lab introduces you to some advanced R functions for plotting data on maps.

Additional resources: You may find the following tutorial useful when completing this lab: "How to Plot and Animate Data on Maps Using the R Programming Language," David J. Lilja, University of Minnesota Digital Conservancy, June, 2021, https://z.umn.edu/mapsUsingR.

What to do:

In this lab, you will again use the house price data set available here:
https://www.kaggle.com/harlfoxem/housesalesprediction

With the column definitions described here:
https://www.kaggle.com/harlfoxem/housesalesprediction/discussion/207885

1. Using the regression model you developed in Problem 1 from Lab 6, compute the percent error for each prediction of the house prices in your testing set. This percent error is simply 100*[(predicted price – actual price) / (actual price)]. Plot the absolute value of this percent error on an appropriate chart. Your chart should be a simple, static map with no animation. Notice that if you plot the errors for every prediction, you will end up with many overlapping dots making it hard to see anything. Instead, plot only a fraction $q$ of the errors. Make separate plots for random samples of the errors with $q$=0.01, 0.05, 0.10, and 1.0. Use the color of the data points to show the percent error for each house's predicted price, with red showing the largest error and green the smallest error. Use the size of each data point to show the corresponding price. Do you see any correlation between the price of a house and the quality of your predictions? For instance, do the predictions tend to be better for houses with higher prices, or vice versa? Or is there no obvious trend?

#####NAKE SURE TO ALSO INSTALL PACKAGE GIFSKI###

2. Animate the plot for the full error list from Problem 1 using the year each house was built as the time parameter for the animation. Set the animation to last 30 seconds.

3. By modifying the zip code segmentation that you developed in Problem 2 of Lab 6, plot the average percent error for each zip code on an appropriate map. Plot the dot for this error for each zip code at the point defined by the mean of all latitudes and the mean of all longitudes within the zip code. Use the color of the data points to show the percent error and the size to show the median price for each zip code. Do you see any correlation between the median prices of houses within a zip code and the quality of your predictions? For instance, do the predictions tend to be better for zip codes with more expensive houses, or vice versa? Or is there no obvious trend?