

PSYC7397 HW4

Instructions for the HW:

- This HW is due on Mar 1 (Friday)
- Do HW in .Rmd file, Render .Rmd to PDF format, and upload to **Github**

Your Name : Blaine Lewis

Problem 1

Follow the instructions to submit problem 2 (a ggplot2 example) and problem 3 (reference, abstract, and keyword) to Github.

- Go to <https://github.com> and Create your GitHub account.
- Send me your GitHub ID to me by email so that I can add you in PSYC7397 repository.
- Watch the following two video tutorial on GitHub
 - RStudio and Git - an Overview (Part 1) to learn about github (<https://www.youtube.com/watch?v=KjLycV1IWqc>)
 - RStudio and Git - an Example (Part 2) to learn about using github within Rstudio (<https://www.youtube.com/watch?v=qcjpHFwCugE>)
- Create a project linked to our PSYC7397 GitHub repository (URL: <https://github.com/sunboklee/PSYC7397>) in a RStudio (Video will tell you how to do that)
- Submit (or Commit in GitHub terminology) your ggplot2 code (problem 2) with your name to PSYC7397_ggplot2_library.Rmd in PSYC7397 repository.
- Submit (or Commit in GitHub terminology) your reference, abstract, and keyword (problem 3) with your name to PSYC7397_ggplot2_library.Rmd in PSYC7397 repository.

Problem 2

Think about a plot that could be useful in your research area, and google how to create the plot using ggplot2. Your code for plotting the plot **should be a working example** including data (probably you need built-in R data: `data()` will give you a list of built-in data). I want you to **explain the details of the code to your friends** in our next class. So be prepared to explain how your code works. **Any plot in our lecture slides are not allowed.** Something new and useful.

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse
## v ggplot2 3.1.0      v purrr   0.2.5
## v tibble  2.0.1      v dplyr  0.7.8
## v tidyr   0.8.2      v stringr 1.3.1
## v readr   1.3.1      v forcats 0.3.0

## -- Conflicts ----- tidyverse
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

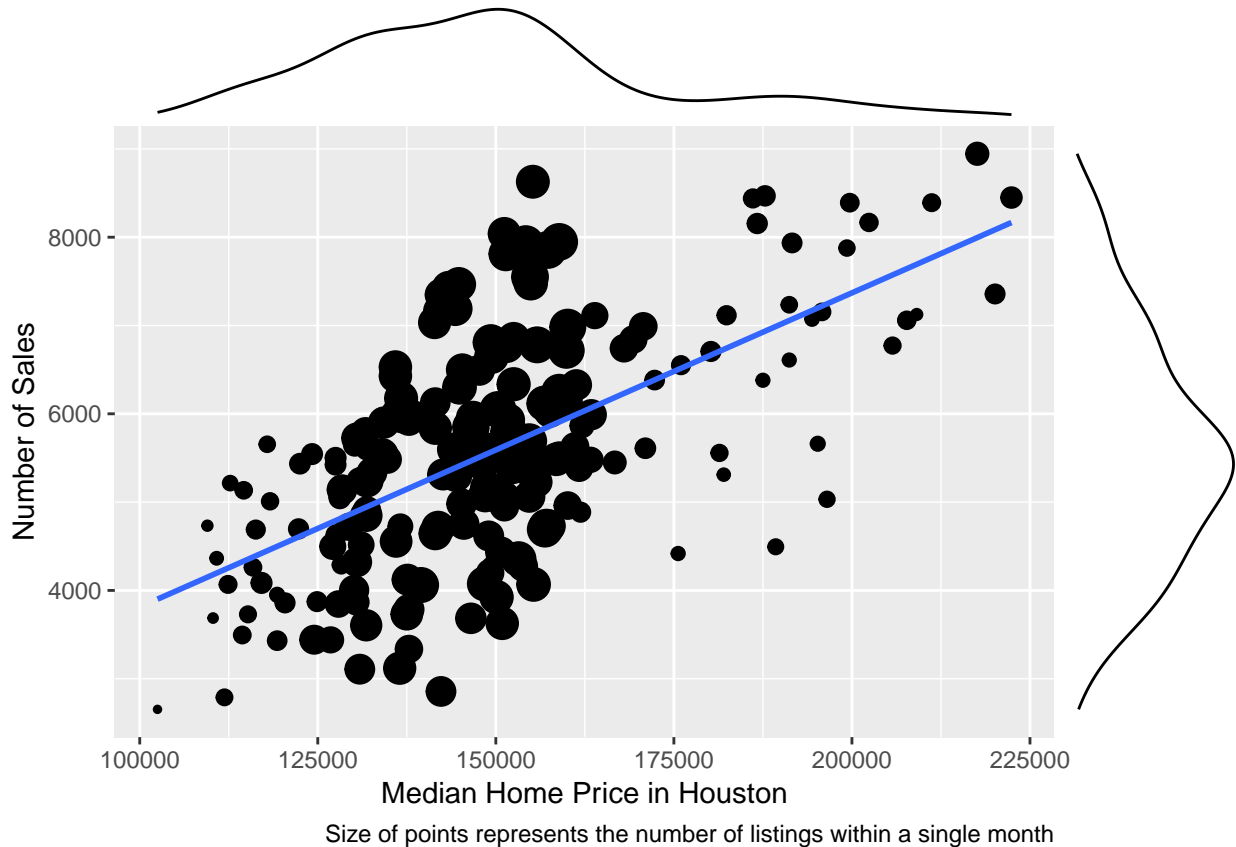
```
library(ggExtra)
```

```
filtered <- txhousing %>%
  filter(city == "Houston")
```

```
scatter <- ggplot(filtered, aes(x = median, y = sales, size = listings)) +
  geom_point() +
  theme(legend.position = "none") +
  labs(x = "Median Home Price in Houston", y = "Number of Sales",
       caption = "Size of points represents the number of listings within a single month") +
  geom_smooth(method = "lm", se = FALSE)

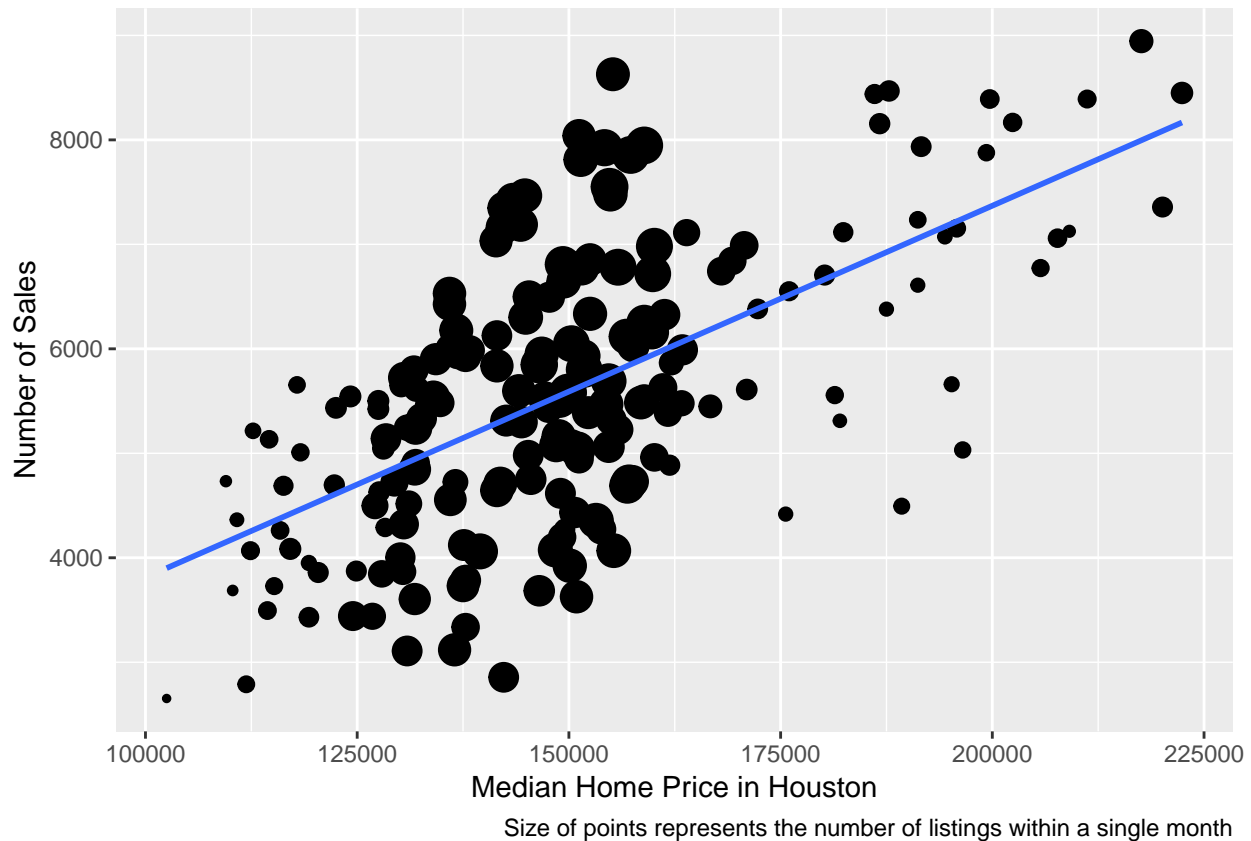
ggMarginal(scatter, type = "density")
```

Warning: Removed 1 rows containing missing values (geom_point).



```
scatter #I am unsure how to suppress the basic scatterplot
```

Warning: Removed 1 rows containing missing values (geom_point).



Problem 3

Find a paper in your research field that used **data mining**. Provide **APA reference** and **abstract** below. I will collect them and will provide a single document that contains the list of references and abstracts, so that we can have a better sense of what kind of data mining research are there. Also be prepared to briefly explain the research you find.

APA Citation Speer, A. B. (2018). Quantifying with words: An investigation of the validity of narrative-derived performance scores. *Personnel Psychology*, 71(3), 299-333.

Abstract Performance appraisal research has focused almost entirely on traditional numerical ratings despite narrative text comments regularly being collected within appraisals. This study investigated the theory and utility of leveraging narrative comments to better understand employee performance. Narrative sentiment scores were derived using text mining on a large sample of narrative comments. These scores were then applied to an independent set of 2 years of performance data. It was assumed that narrative comments would reflect true performance variance that overlaps with traditional ratings, but also that they would capture incremental variance due to increases in total information and a reduction in rater-motivated biases in contexts in which narrative data were not explicitly linked to administrative outcomes. The derived narrative scores were reliable across years, converged with traditional numerical ratings and explained incremental variance in future performance outcomes (performance ratings, involuntary turnover, promotions, and pay increases). Collectively, this study highlights how narratives can enhance performance measurement and demonstrates how these data can be economically scored in applied settings.