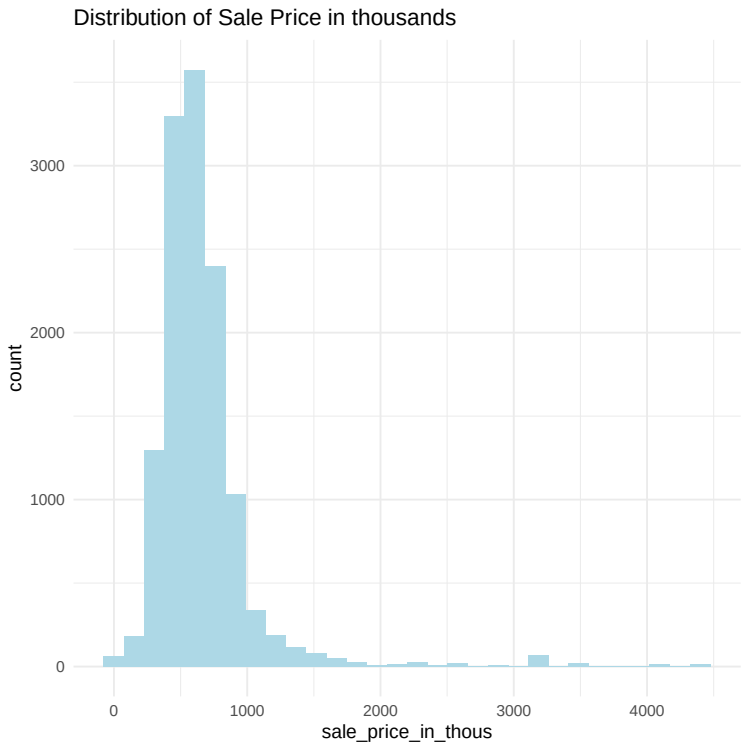
The results below are generated from an R script.

```
# Assignment: ASSIGNMENT 4
# Name: Reppeto, Brian
# Date: 2023-06-27

## Load the ggplot2 package

library(tidyverse)
library(readxl)
library(ggplot2)
library(dplyr)
library(conflicted)
#library(plyr)
theme_set(theme_minimal())

## Set the working directory to the root of your DSC 520 directory
setwd("~/DSC520/Week_4")

## Load the `data` to
housing_df <- read_xlsx("week-6-housing.xlsx")

#housing_df %>% rename("Sale_Price" = "Sale Price")

#sum_by_day <-apply(housing_df$'Sale Price',housing_df$'Sale Date',sum)


## 1.
ttl_house_prices <- apply(housing_df[c('Sale Price')], 2, sum)

print (ttl_house_prices)

## Sale Price
## 8500391149

## 2.
ttl_sales_by_zip <- aggregate(`Sale Price` ~  zip5, data = housing_df, mean)

print (ttl_sales_by_zip)

##    zip5 Sale Price
## 1 98052   649375.4
## 2 98053   672623.7
## 3 98059   645000.0
## 4 98074   951543.8

## 3.

housing <-
  rename(housing_df, sale_price = `Sale Price`, sale_date = `Sale Date`)

housing <- mutate(housing, month = lubridate :: month (sale_date),
                           year = lubridate :: year (sale_date))

sales_year <- aggregate(sale_price ~ year, data = housing, sum)
```

```r
housing <- mutate(housing, sale_price_in_thous = sale_price / 1000)

print(sales_year)
```

```
##    year sale_price
## 1  2006  919005546
## 2  2007  850285247
## 3  2008  755045696
## 4  2009  454417263
## 5  2010  592828305
## 6  2011  693256612
## 7  2012  747585171
## 8  2013  860712529
## 9  2014  792182425
## 10 2015  915474082
## 11 2016  919598273
```

```r
## 4
ggplot(housing, aes(x = sale_price_in_thous)) +
  geom_histogram(fill = "lightblue", bins = 30) +
  labs(title = "Distribution of Sale Price in thousands")
```




```r
## 5.
## There are outliers in the data for the sale_price variable as shown in the
## graph above.

## 6.
housing <- mutate(housing, price_per_sf =  sale_price/square_feet_total_living)
```

```r
housing <- mutate(housing, commision = .06)

housing <-
  mutate(housing, net_sales_price = (sale_price - (sale_price * commision)))

#head(select(housing,price_per_sf))

#head(select(housing, net_sales_price, commision, sale_price_in_thous))

##view(housing)
```

The R session information (including the OS info, R version and all packages used):

```r
sessionInfo()

## R version 4.3.0 (2023-04-21)
## Platform: aarch64-apple-darwin20 (64-bit)
## Running under: macOS Ventura 13.4.1
##
## Matrix products: default
## BLAS:   /System/Library/Frameworks/Accelerate.framework/Versions/A/Frameworks/vecLib.framework/Versio
## LAPACK: /Library/Frameworks/R.framework/Versions/4.3-arm64/Resources/lib/libRlapack.dylib;  LAPACK ve
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## time zone: America/New_York
## tzcode source: internal
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] conflicted_1.2.0 readxl_1.4.2     lubridate_1.9.2  forcats_1.0.0    stringr_1.5.0
##  [6] dplyr_1.1.2      purrr_1.0.1      readr_2.1.4      tidyr_1.3.0      tibble_3.2.1
## [11] ggplot2_3.4.2    tidyverse_2.0.0
##
## loaded via a namespace (and not attached):
##  [1] gtable_0.3.3     highr_0.10       compiler_4.3.0   tidyselect_1.2.0 scales_1.2.1
##  [6] fastmap_1.1.1    R6_2.5.1         labeling_0.4.2   generics_0.1.3   knitr_1.43
## [11] munsell_0.5.0    pillar_1.9.0     tzdb_0.4.0       rlang_1.1.1      utf8_1.2.3
## [16] cachem_1.0.8     stringi_1.7.12   xfun_0.39        timechange_0.2.0 memoise_2.0.1
## [21] cli_3.6.1        withr_2.5.0      magrittr_2.0.3   grid_4.3.0       rstudioapi_0.14
## [26] hms_1.1.3        lifecycle_1.0.3  vctrs_0.6.3      evaluate_0.21    glue_1.6.2
## [31] farver_2.1.1     cellranger_1.1.0 fansi_1.0.4      colorspace_2.1-0 tools_4.3.0
## [36] pkgconfig_2.0.3

Sys.time()

## [1] "2023-07-02 18:49:37 EDT"
```