

# Assignment 2

## Deconstruct, Reconstruct Web Report

Timm Rahrt



### Assessment declaration checklist

Please carefully read the statements below and check each box if you agree with the declaration. If you do not check all boxes, your assignment will not be marked. If you make a false declaration on any of these points, you may be investigated for academic misconduct. Students found to have breached academic integrity may receive official warnings and/or serious academic penalties. Please read more about academic integrity here (<https://www.rmit.edu.au/students/student-essentials/assessment-and-exams/academic-integrity>). If you are unsure about any of these points or feel your assessment might breach academic integrity, please contact your course coordinator for support. It is important that you DO NOT submit any assessment until you can complete the declaration truthfully.

#### By checking the boxes below, I declare the following:

- ☒ I have not impersonated, or allowed myself to be impersonated by, any person for the purposes of this assessment
- ☒ This assessment is my original work and no part of it has been copied from any other source except where due acknowledgement is made.
- ☒ No part of this assessment has been written for me by any other person except where such collaboration has been authorised by the lecturer/teacher concerned.
- ☒ Where this work is being submitted for individual assessment, I declare that it is my original work and that no part has been contributed by, produced by or in conjunction with another student.
- ☒ I give permission for my assessment response to be reproduced, communicated compared and archived for the purposes of detecting plagiarism.
- ☒ I give permission for a copy of my assessment to be retained by the university for review and comparison, including review by external examiners.

#### I understand that:

- ☒ Plagiarism is the presentation of the work, idea or creation of another person as though it is your own. It is a form of cheating and is a very serious academic offence that may lead to exclusion from the University. Plagiarised material can be drawn from, and presented in, written, graphic and visual form, including electronic data and oral presentations. Plagiarism occurs when the origin of the material used is not appropriately cited.
- ☒ Plagiarism includes the act of assisting or allowing another person to plagiarise or to copy my work.

#### I agree and acknowledge that:

- # Deconstruct

# Original

# HOME PRICE-TO-INCOME RATIOS OF LARGE U.S. CITIES

Large cities have 1 million or more residents.

Home price-to-income ratio was calculated by dividing median home price by the median annual household income.

HOME PRICE-TO-INCOME RATIO

2 4 6 8 10 12

| City              | Ratio |
|-------------------|-------|
| Los Angeles, CA   | 12.5  |
| San Jose, CA      | 10.5  |
| Long Beach, CA    | 10.3  |
| San Diego, CA     | 9.9   |
| New York, NY      | 9.8   |
| San Francisco, CA | 9.2   |
| San Jose, CA      | 9.0   |
| San Francisco, CA | 8.4   |
| San Francisco, CA | 8.3   |
| San Francisco, CA | 7.3   |
| San Francisco, CA | 6.5   |
| San Francisco, CA | 6.4   |
| San Francisco, CA | 6.4   |
| San Francisco, CA | 6.0   |
| San Francisco, CA | 0.9   |
| San Francisco, CA | 0.9   |
| San Francisco, CA | 5.9   |
| San Francisco, CA | 5.9   |
| San Francisco, CA | 5.9   |
| San Francisco, CA | 5.8   |
| San Francisco, CA | 5.8   |
| San Francisco, CA | 5.7   |
| San Francisco, CA | 5.6   |
| San Francisco, CA | 5.5   |
| San Francisco, CA | 5.3   |
| San Francisco, CA | 5.3   |
| San Francisco, CA | 5.3   |

\*City's metropolitan area shaded according to data for visibility.

Source: Construction Coverage analysis of Zillow and U.S. Census Bureau data

Source: *Construction Coverage analysis of Zillow and U.S. Census Bureau data (2024)*  
<https://www.visualcapitalist.com/mapped-home-price-to-income-ratio-of-large-u-s-cities/>  
(<https://www.visualcapitalist.com/mapped-home-price-to-income-ratio-of-large-u-s-cities/>)

## Objective and Audience

The objective and audience of the original data visualisation chosen can be summarised as follows:

### Objective

The original visualisation, featured on **Visual Capitalist**, utilises data from **Construction Coverage analysis of Zillow and U.S. Census Bureau**, aims to depict housing affordability across major U.S. cities through a comparison of home price-to-income ratio's. The plot employs a colour gradient from blue to red, where blue represents more affordable cities with lower ratio's and red indicates less affordable cities with higher ratio's. The visualisation is organised in a radial layout, where the least affordable cities are at the top and the most affordable at the bottom. Furthermore, a map of the U.S. highlights each city's location as well as their respective metropolitan area.

This graphic is designed with the following key objectives in mind:

1. It illustrates affordability variations across large U.S. cities based on home price-to-income ratios through colour coding to group similar priced ones together and display relationships at one glance.
2. The plot uses a visual ranking of cities which serves as an affordability spectrum. The ranking is ordered from least to most affordable and the text elements follow each bar plot rotation to emphasise which textual label belongs to which bar.
3. The visualisation implements a geographic context map to identify and compare patterns and groups of less or more affordable metropolitan areas in the U.S, while using transparency of colour to highlight the 'Home Price-to-Income Ratio' intensity.

### Audience

The visualisation's target audience include a lot of different interest groups/stakeholders such as:

1. **Real estate professionals** who rely on housing affordability data to guide their clients on investment opportunities using data-driven decision making. This visualisation helps them to identify cities which are too expensive for consideration or areas which might develop promising overtime or due to its proximity to more expensive areas.
2. **The general public** who is interested in understanding housing affordability across major U.S. cities to support their relocating or financial plannings as they can see on one glance which areas might be out of their budget. The visualisation also allows for follow-up questions such as "Why is this city ranking so poorly in the Price-to-Income Ratio?".
3. **Urban planners** who need data to identify potential cities to invest in or cities with challenges. As mentioned before, they can also use this visualisation just like Real estate professionals to identify promising areas or so called 'Rising Stars'.
4. **Policymakers**, who focus on reducing economic disparities within similar regions, can use this plot to also ask themselves which factors are causing the 'Home Price-to-Income-Index' to decline or increase. They can also find patterns or similarities of low income cities and scrape for reasons on

how new policies might increase the economical power of such region to support locals.

The visualisation does contain multiple audience issues which we will need to address.

## Critique

The visualisation chosen had the following three main issues:

### 1. ***Unclear Communication of the Visualization's Objective***

The visualisation doesn't answer a clear 'So what' question and it doesn't clarify its clear purpose. A clear practical question would be "Which cities are at risk of becoming unaffordable" or "How does affordability compare in the U.S.?". Therefore, the title of the plot doesn't fulfill its purpose leaving the audience clueless of what the Visualisation or Data is actually trying to achieve (Trifecta Layer Q has been ignored (Baglin, 2023)). Lastly, the clear focus on aesthetics detracts from the delivery of an actionable insight which "deceives the receiver" (Kirk, 2012) and breaks one of Andy Kirk (2012) guiding principles.

Additionally, the subtitle doesn't add any real value to the visualisation's objective as 54 cities have over 1 million residents. It would be more effective to group the cities into regions to more precisely understand which part of the U.S. might offer a promising 'Home Price-to-Income Ratio'.

This problem of an unclear objective can be solved by using a title and subtitle that gives the visualisation a clear purpose, so that the audience understands the effect the plot will have and how it will increase their knowledge.

### 2. ***Radial Layout and Overcrowding (Clutter)***

The visualisation clearly emphasises form (style) over function (substance). While the useage of a radial layout is visually engaging, it complicates comparisons of "Home Price-to-Income Ratio's" across U.S. cities. The curvature can complicate to interpret the actual meaning of the labels (or data) and the positioning is very disturbing for the audience. Additionally, the author uses uneven spacing of cities with 'similar affordability ratios', which also distorts comparisons.

Secondly, the visualisation is using text elements ineffectively losing readability and focus. The textual elements are not serving its main purpose of enhancing the story of the visualisation's objective.

Improvement suggestions would be to use a hierarchical and horizontal textual layout for more intuitive comparisons (Baglin, 2023), as vertical or rotating text is hard to read. The author also needs to use labels sparingly to increase their meaning and gain focus on the important aspects of the visualisation. Lastly, it is mandatory to use consistent scaling and spacing to ensure a proportional and visually pleasing representation of the data.

### 3. ***Ineffective Use of Colour and Gestalt principles***

The visualisation makes excessive use of colour when attempting to represent the affordability spectrum of U.S. cities, while also failing to group or emphasise meaningful patterns. This overuse of colour diminishes its effectiveness to highlight or draw attention to key data elements which could support the audience's interpretation of the data.

The colour palette is lacking accessibility, as individuals with colour blindness, who can't distinguish between shades of blue and red, are excluded from interpreting the data. Furthermore, the colour palette isn't suitable for a black and white print, as the colour intensity varies from dark (red) to light (red/blue) and then back to dark (blue).

Lastly, the colour of textual map elements do not sufficiently contrast with the background, which further reduces readability.

An improvement suggestion is to reduce reliance on colour by using Gestalt principles like proximity or similarity to highlight similar data values without bombarding the audience with colour. This would also allow using colour strategically to highlight important data insights and to enhance the visualisation's objective. Moreover, it is crucial to improve accessibility by using a colour-blind friendly palette that also gradients from light (affordable) to dark (expensive), ensuring compatibility with black and white prints.

# Reconstruct

## Code

The following code was used to fix the issues identified in the original.

```
library(ggplot2) # For visualisation
library(dplyr) # For data preprocessing
library(readr) # For data import
library(scales) # For axis customisation
library(rvest) # For webscraping
# Webscrape the data - Save the url
url <- "https://www.visualcapitalist.com/mapped-home-price-to-income-ratio-of-large-u-s-cities/"
url_page <- read_html(url)
# Extract and clean the table
df <- url_page %>%
  html_element("table") %>%
  html_table(trim = TRUE)

# Step 1: Create a DataFrame of unique States to later map a new Region column to df
state_ <- data_frame(State = c("CA", "NY", "FL", "MA", "WA", "OR", "CO", "AZ", "DC", "TX", "TN", "NV", "NC",
                              "NM", "VA", "GA", "LA", "MN", "IL", "NE", "OH", "WI", "PA", "KY", "MO",
                              "IN",
                              "OK", "MD", "KS", "MI"),
  Region = c(
    "Pacific West", "Northeast", "South", "Northeast", "Pacific West", "Pacific West",
    "Mountain West", "Mountain West", "Northeast", "South Central", "South", "Mountain West",
    "South", "Mountain West", "South", "South", "South Central", "Midwest", "Midwest")
```

```

    "Midwest", "Midwest", "Midwest", "Northeast", "South", "Midwest", "Midwest",
    "South Central", "Northeast", "Midwest", "Midwest"))

# Step 2: First remove special characters like '$' then join the newly created Data
Frame to our existing one `df`
cleaned_df <- df %>% mutate(`Median home price` = as.numeric(gsub("[$,]", "", `Medi
an home price`)),
                           `Median household income` = as.numeric(gsub("[$,]", "",
`Median household income`))) %>%
  left_join(state_, by = "State")

# Step 3: Create a second new variable which groups Price-to-income into five categ
ories
# We need to factor the newly created values, also to ensure R doesn't sort the val
ues alphabetically
cleaned_df <- cleaned_df %>% mutate(Affordability = factor(case_when(`Price-to-inco
me` <= 3 ~ "Very high",
                                                                    `Price-to-income` > 3
& `Price-to-income` <= 5 ~ "High",
                                                                    `Price-to-income` > 5
& `Price-to-income` <= 7 ~ "Medium",
                                                                    `Price-to-income` > 7
& `Price-to-income` <= 9 ~ "Low",
                                                                    TRUE ~ "Very low"),
                           levels = c("Very high", "High", "Medium", "Lo
w", "Very low")))

# Step 4: Plot the scatter plot
p1 <- ggplot(cleaned_df, aes(x=`Median household income`, y=`Median home price`, co
lor = Affordability)) +
  geom_point(size = 3, alpha = 0.8) + # Create a scatter plot with increased point
size and transparency
  scale_x_continuous(labels = dollar_format(scale = 1e-3, suffix = "k"), breaks = s
eq(0, 200000, 25000)) + # Trim the x-label
  scale_y_continuous(labels = dollar_format(scale = 1e-3, suffix = "k"), breaks = s
eq(0, 1500000, 250000)) + # Trim the y-label
  scale_color_manual(values = RColorBrewer::brewer.pal(5, "PuBu")) + # Use a colorb
lind-friendly palette
  facet_wrap(~ Region, ncol = 3) + # Create a 2x3 scatter plot figure split by Regi
on
  labs( # Add text elements to the plot
    title = "How does Home Price-to-Income compare in the U.S.?",
    subtitle = "Relationship between Median Home Price and Household Income by Regi
on",
    x = "Median Household Income (USD)",
    y = "Median Home Income (USD)",
    color = "Affordability:",
    caption = "Price-to-Income Ratio: \nVery High: ≤ 3 | High: 3-5 | Medium: 5-7 |
Low: 7-9 | Very Low: >9"
  ) + theme_minimal() + # Add a theme for clear and simple plots
  theme(strip.text = element_text(size = 9, face = "bold"), # Customise and positio

```

*n the text elements*

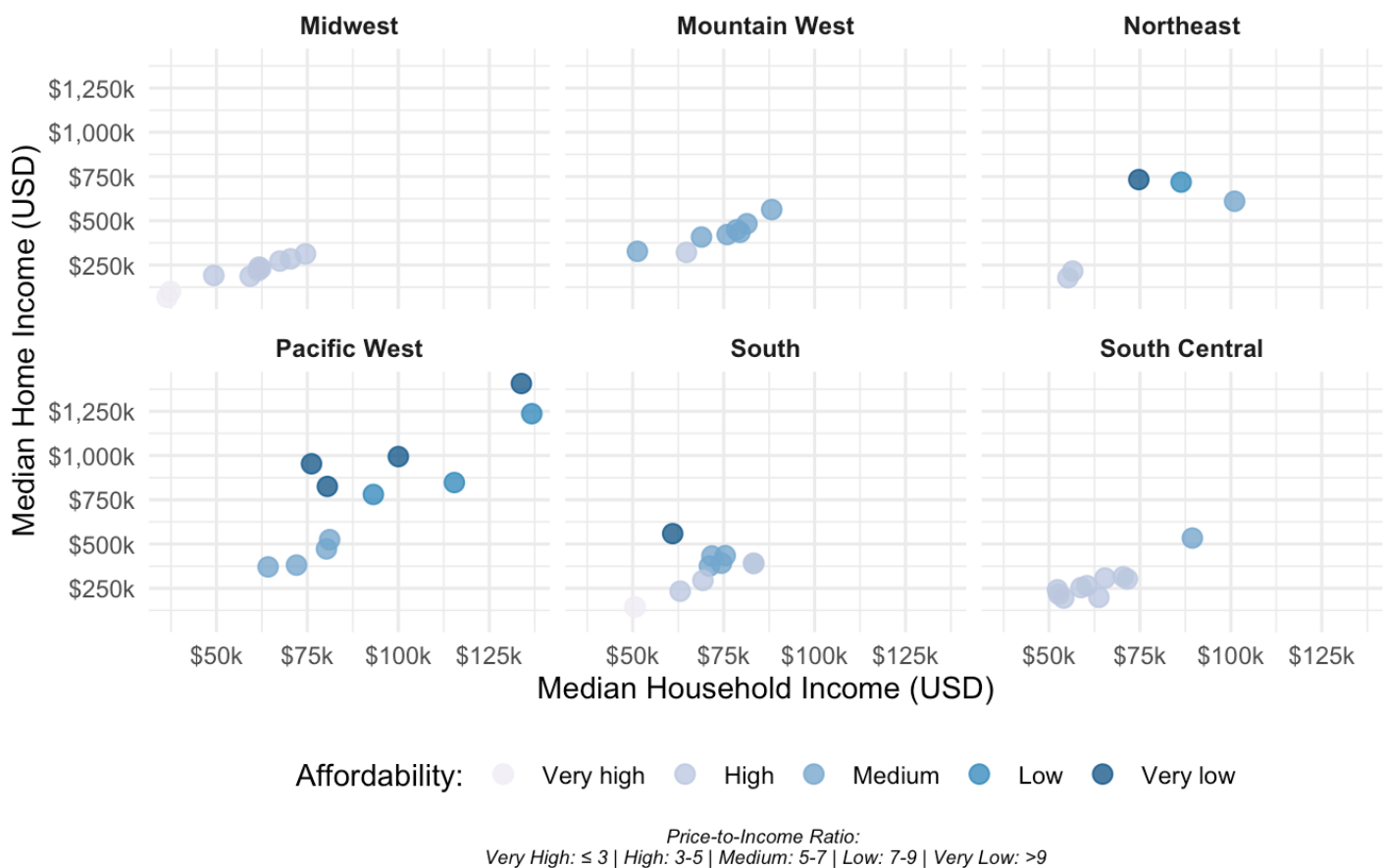
```
legend.position = "bottom",
plot.title = element_text(size = 16, face = "bold", hjust = 0),
plot.subtitle = element_text(size = 12, hjust = 0),
plot.caption = element_text(size = 7, face = "italic", hjust = 0.5))
```

## Reconstruction

The following plot fixes the main issues in the original.

### How does Home Price-to-Income compare in the U.S.?

Relationship between Median Home Price and Household Income by Region



## References

The reference to the original data visualisation choose, the data source(s) used for the reconstruction and any other sources used for this assignment are as follows:

- Visual Capitalist. (2024). *Mapped: Home price-to-income ratio of large U.S. cities*. Retrieved November 07, 2024, from Visual Capitalist website: <https://www.visualcapitalist.com/mapped-home-price-to-income-ratio-of-large-u-s-cities/> (<https://www.visualcapitalist.com/mapped-home-price-to-income-ratio-of-large-u-s-cities/>)
- Baglin, J. (2023). *Data visualisation: From theory to practice*. Retrieved November 17, 2024, from [http://www.dark-star-161610.appspot.com/secured/\\_book/design-and-integrity.html](http://www.dark-star-161610.appspot.com/secured/_book/design-and-integrity.html) ([http://www.dark-star-161610.appspot.com/secured/\\_book/design-and-integrity.html](http://www.dark-star-161610.appspot.com/secured/_book/design-and-integrity.html))

- CRAN. (2024). *urltools: Tools for URL handling and parsing*. Retrieved November 12, 2024, from CRAN website: <https://cran.r-project.org/web/packages/urltools/vignettes/urltools.html> (<https://cran.r-project.org/web/packages/urltools/vignettes/urltools.html>)
- RDocumentation. (2024). *dollar\_format function*. Retrieved November 14, 2024, from RDocumentation website: [https://www.rdocumentation.org/packages/scales/versions/0.4.1/topics/dollar\\_format](https://www.rdocumentation.org/packages/scales/versions/0.4.1/topics/dollar_format) ([https://www.rdocumentation.org/packages/scales/versions/0.4.1/topics/dollar\\_format](https://www.rdocumentation.org/packages/scales/versions/0.4.1/topics/dollar_format))
- GeeksforGeeks. (2024). *How to add caption to a ggplot in R*. Retrieved November 15, 2024, from GeeksforGeeks website: <https://www.geeksforgeeks.org/how-to-add-caption-to-a-ggplot-in-r/> (<https://www.geeksforgeeks.org/how-to-add-caption-to-a-ggplot-in-r/>)