Real_Estate_Data

Table of contents

Libraries

```
library(sf)
```

```
Linking to GEOS 3.13.1, GDAL 3.11.0, PROJ 9.6.0; sf_use_s2() is TRUE
```

```
library(dplyr)
```

```

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```
library(tidyr)
library(scales)
```

Load Map Data

First, we load the data map from the `.shp` file and get the corresponding geometries for the townships in Taiwan. For the scope of this project, we are only studying the data from Hsinchu County and City as well as Taipei City.

```
# Load township-level shapefile
map_data <- st_read("coords/TOWN_MOI_1140318.shp")
```

```
Reading layer `TOWN_MOI_1140318' from data source
  `C:\Users\samue\OneDrive\Documents\NTHU\1s_fall_25-26\PBA\Group_project\1022\aging_medical_
  using driver `ESRI Shapefile'
Simple feature collection with 368 features and 7 fields
Geometry type: MULTIPOLYGON
Dimension:     XY
Bounding box:  xmin: 114.3593 ymin: 10.37135 xmax: 124.5612 ymax: 26.38528
Geodetic CRS:  GCS_TWD97[2020]
```

Load Dataset

The data set can is managed and accessed through the load_data script. This loads all of the data set necessary for this visualization. Afterwards, we view the `manifest.csv` to get a list of all the files from the Real Estate data set and their descriptions.

```
suppressMessages(
  suppressWarnings(
    invisible(capture.output(source("load_data.R")))
  )
)

# Look at the Metadata
manifest_list <- real_estate_data[["manifest.csv"]]
manifest_list
```

```
# A tibble: 231 x 3
   name                   schema               description
   <chr>                  <chr>                <chr>
 1 a_lvr_land_a.csv       schema-main.csv
 2 a_lvr_land_a_build.csv schema-build.csv
 3 a_lvr_land_a_land.csv  schema-land.csv
 4 a_lvr_land_a_park.csv  schema-park.csv
 5 a_lvr_land_b.csv       schema-main-sale.csv
 6 a_lvr_land_b_land.csv  schema-land.csv
 7 a_lvr_land_b_park.csv  schema-park.csv
 8 a_lvr_land_c.csv       schema-main-rent.csv
 9 a_lvr_land_c_build.csv schema-build.csv
10 a_lvr_land_c_land.csv  schema-land.csv
# i 221 more rows
```

Average Land Prices

Hsinchu

In this section, we will try to analyze the elderly population data from Hsinchu City nd County and see its relationship with the average land prices.

Filter Hsinchu Data

Here we get the real estate data of Hsinchu get only the columns `TOWNNAMES` as well their average land prices by getting the mean of all records per Township. Since Hsinchu City in the data doesn't have a record per district, we will get the data of the entire city and split them into three districts having the equal land value as the city itself.

```
hsinchu <- list("j_lvr_land_a.csv", "o_lvr_land_a.csv") %>%
  lapply(function(file) {
    real_estate_data[[file]] %>%
      {
        colnames(.) <- as.character(unlist(.[1, ]))
        .[-1, ]
      } %>%
      select(1, 13, 23)
  }) %>%
  bind_rows()


hsinchu <- hsinchu %>%
  filter(`main use` %in% c("  ", "  ", "  ")) %>%
  rename(TOWNNAME = `The villages and towns urban district`) %>%
  rename(LANDPRICE = `the unit price (NTD / square meter)`)


hsinchu_ave_price <- hsinchu %>%
  group_by(TOWNNAME) %>%
  summarise(
    avg_price = mean(
      as.numeric(str_replace_all(LANDPRICE, ",", "")),
      na.rm = TRUE
    )
  )

price_to_copy <- hsinchu_ave_price %>%
  filter(TOWNNAME == "  ") %>%
  pull(avg_price)

hsinchu_ave_price <- hsinchu_ave_price %>%
  add_row(
    TOWNNAME = c(" ", "  ", " "),
    avg_price = price_to_copy
  ) %>%
```

```
  filter(TOWNNAME != "  ")

hsinchu_ave_price
```

```
# A tibble: 14 x 2
   TOWNNAME avg_price
   <chr>        <dbl>
 1         54250.
 2        108159.
 3         24226.
 4        114650.
 5         85464.
 6         67473
 7         71223.
 8        165146.
 9         73184.
10        102304.
11         48663.
12          117351.
13        117351.
14          117351.
```

Get Hsinchu Map Data

```
hsinchu_map_data <- map_data %>%
  filter(COUNTYNAME %in% c("  ", "  ")) %>%
  select(3, 4, 8)
```

Merge Map and Average price

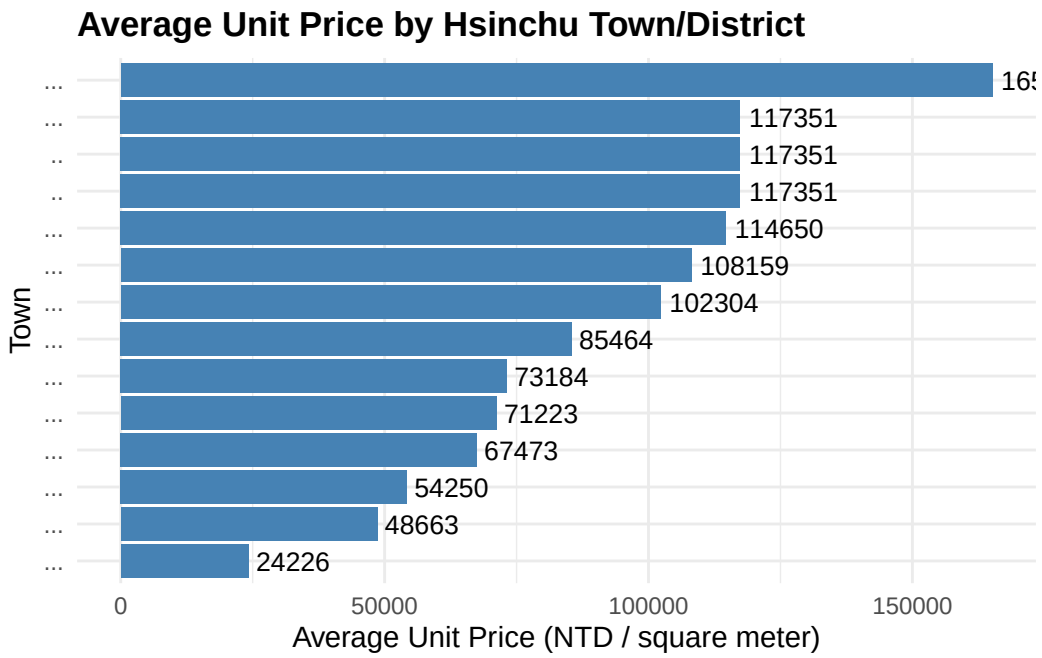
After having the real estate data and their corresponding map coordinates. We'll merge the two tibbles together.

```
hsinchu_map_data <- hsinchu_map_data %>%
  left_join(hsinchu_ave_price, by = "TOWNNAME") %>%
  replace_na(list(avg_price = NA))
```

Bar Plot of Averge Land Prices in Hsinchu

Here we will generate a plot of Hsinchu townships and districts vs their average land prices per square meter in TWD.

```
ggplot(hsinchu_ave_price, aes(x = reorder(TOWNNAME, avg_price), y = avg_price)) +
  geom_col(fill = "steelblue") +
  geom_text(aes(label = round(avg_price)),
            hjust = -0.1,
            size = 3.5) +
  coord_flip() +
  labs(
    x = "Town",
    y = "Average Unit Price (NTD / square meter)",
    title = "Average Unit Price by Hsinchu Town/District"
  ) +
  theme_minimal() +
  theme(plot.title = element_text(face = "bold"))
```

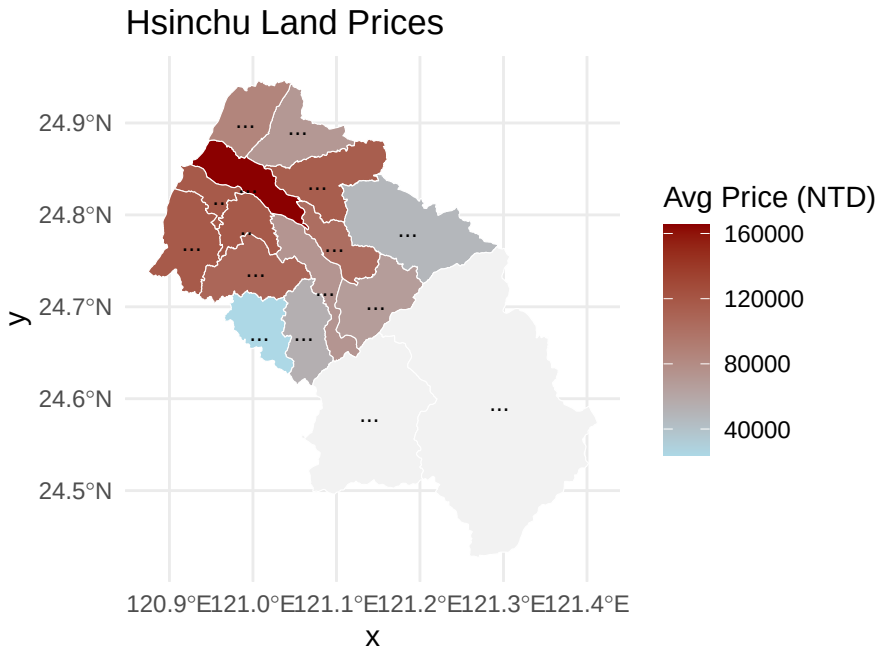


Heatmap of Average Land Prices in Hsinchu Plot

This is also another visualization of Hsinchu land prices in heatmap form. Each color corresponds to the price.

```
ggplot(hsinchu_map_data) +
  geom_sf(aes(fill = avg_price), color = "white") +
  geom_sf_text(aes(label = TOWNNAME), size = 3, color = "black") +
  scale_fill_gradient(low = "lightblue", high = "darkred", na.value = "gray95") +
  labs(
    title = "Hsinchu Land Prices",
    fill = "Avg Price (NTD)"
  ) +
  theme_minimal()
```

```
Warning in st_point_on_surface.sfc(sf::st_zm(x)): st_point_on_surface may not
give correct results for longitude/latitude data
```




```
ggsave("hsinchu_map_large.png", plot = last_plot(), width = 13, height = 8, dpi = 300)
```

```
Warning in st_point_on_surface.sfc(sf::st_zm(x)): st_point_on_surface may not
give correct results for longitude/latitude data
```

Taipei City

In this section, we will now do the same process but in Taipei City.

Filter Taipei Data

```
taipei <- ("a_lvr_land_a.csv") %>%
  lapply(function(file) {
    real_estate_data[[file]] %>%
      {
        colnames(.) <- as.character(unlist(.[1, ]))
        .[-1, ]
      } %>%
      select(1, 13, 23)
  }) %>%
  bind_rows()


taipei <- taipei %>%
  filter(`main use` %in% c("  ", "  ", "  ")) %>%
  rename(TOWNNAME = `The villages and towns urban district`) %>%
  rename(LANDPRICE = `the unit price (NTD / square meter)`)


taipei_ave_price <- taipei %>%
  group_by(TOWNNAME) %>%
  summarise(
    avg_price = mean(
      as.numeric(str_replace_all(LANDPRICE, ",", "")),
      na.rm = TRUE
    )
  )

taipei_ave_price
```

```
# A tibble: 12 x 2
   TOWNNAME avg_price
   <chr>        <dbl>
 1        254165.
 2        298650.
```

```
 3        255615.
 4        211701.
 5        206480.
 6        251513.
 7        224440.
 8        271485.
 9        350564.
10        181216.
11        313034.
12        184172.
```

Get Taipei Map Data

```
taipei_map_data <- map_data %>%
  filter(COUNTYNAME %in% c("  ")) %>%
  select(3, 4, 8)
```

Merge Map and Average price

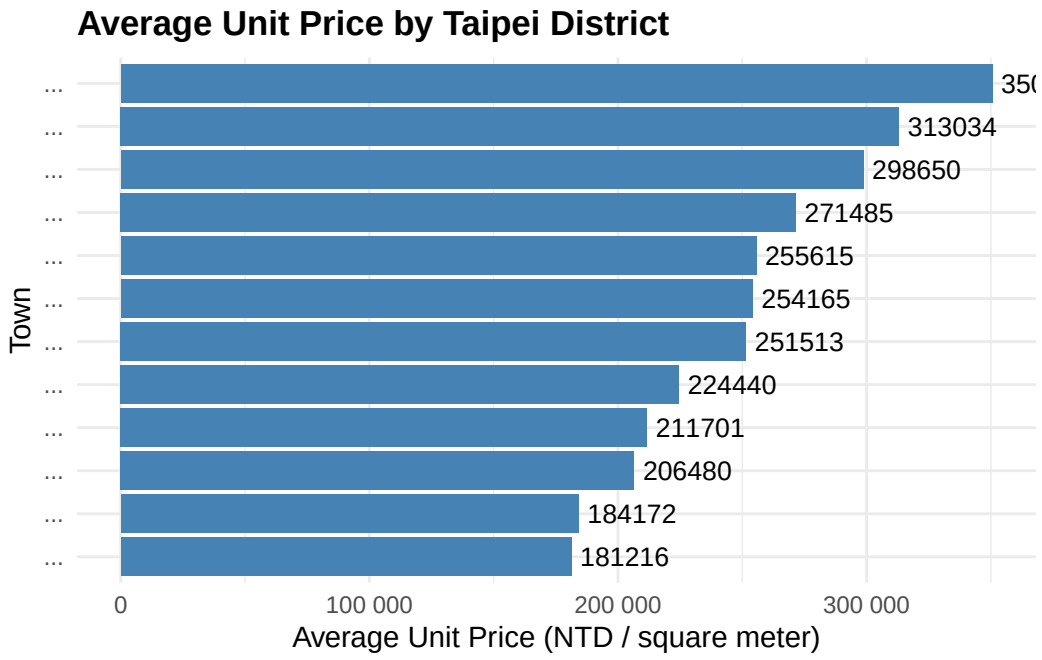
```
taipei_map_data <- taipei_map_data %>%
  left_join(taipei_ave_price, by = "TOWNNAME") %>%
  replace_na(list(avg_price = NA))
```

Bar Plot of Averge Land Prices in Taipei City

The bar plot below shows the average land price of Taipei City per districts.

```
ggplot(taipei_ave_price, aes(x = reorder(TOWNNAME, avg_price), y = avg_price)) +
  geom_col(fill = "steelblue") +
  geom_text(aes(label = round(avg_price)),
            hjust = -0.1,
            size = 3.5) +
  coord_flip() +
  scale_y_continuous(labels = label_number(accuracy = 1)) +
  labs(
    x = "Town",
    y = "Average Unit Price (NTD / square meter)",
    title = "Average Unit Price by Taipei District"
```

```
  ) +
  theme_minimal() +
  theme(plot.title = element_text(face = "bold"))
```

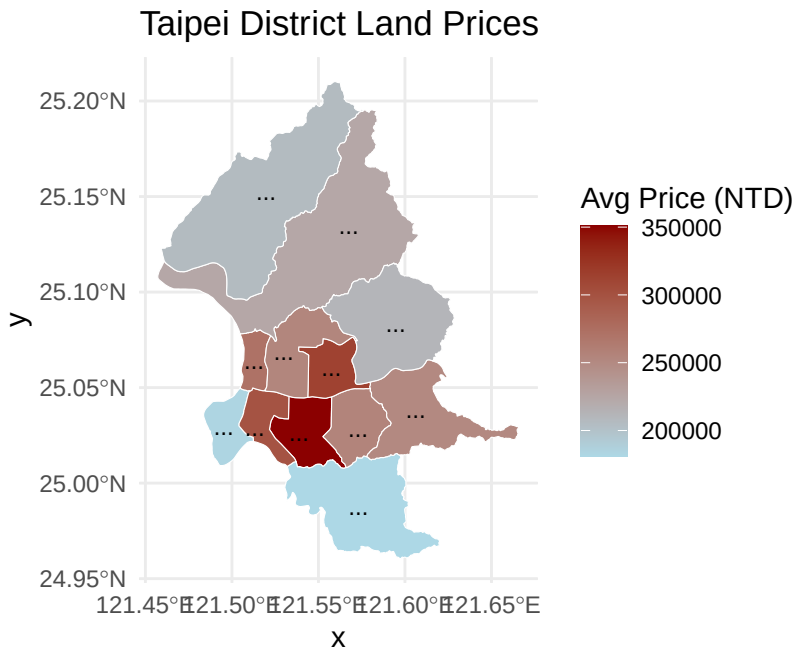


Heatmap of Average Land Prices in Taipei City Plot

Below is the heatmap of Taipei city districts and their corresponding labd prices.

```
ggplot(taipei_map_data) +
  geom_sf(aes(fill = avg_price), color = "white") +
  geom_sf_text(aes(label = TOWNNAME), size = 3, color = "black") +
  scale_fill_gradient(low = "lightblue", high = "darkred", na.value = "gray95") +
  labs(
    title = "Taipei District Land Prices",
    fill = "Avg Price (NTD)"
  ) +
  theme_minimal()
```

```
Warning in st_point_on_surface.sfc(sf::st_zm(x)): st_point_on_surface may not
give correct results for longitude/latitude data
```



Elderly Dependency Ratio and Average Lad Prices

In this section, we will now try to see the elderly population and its relationship with the average land prices per townships and districts.

Hsinchu

Here we try to get the data this time the `aging_data` from a different data set. In this part, we will try to extract the data and merge it from the tibble we made from previous stages.

```
hsinchu_pop_data <- aging_data %>%
  filter(COUNTY %in% c("  ", "  ")) %>%
  select(2, 4, 10) %>%
  mutate(A65UP_A15A64_RAT = as.integer(A65UP_A15A64_RAT)) %>%
  rename(TOWNNAME = TOWN)

hsinchu_pop_data
```

```
# A tibble: 16 x 3
   COUNTY TOWNNAME A65UP_A15A64_RAT
```

```
   <chr>  <chr>                <int>
 1                            13
 2                            23
 3                            31
 4                            34
 5                            20
 6                            19
 7                            28
 8                            38
 9                            36
10                            25
11                            47
12                            15
13                            21
14                             21
15                             24
16                            21
```

```
hsinchu_map_data <- hsinchu_map_data %>%
  left_join(hsinchu_pop_data, by = "TOWNNAME")


hsinchu_map_data$avg_price <- as.numeric(hsinchu_map_data$avg_price)
hsinchu_map_data$A65UP_A15A64_RAT <- as.numeric(hsinchu_map_data$A65UP_A15A64_RAT)
```

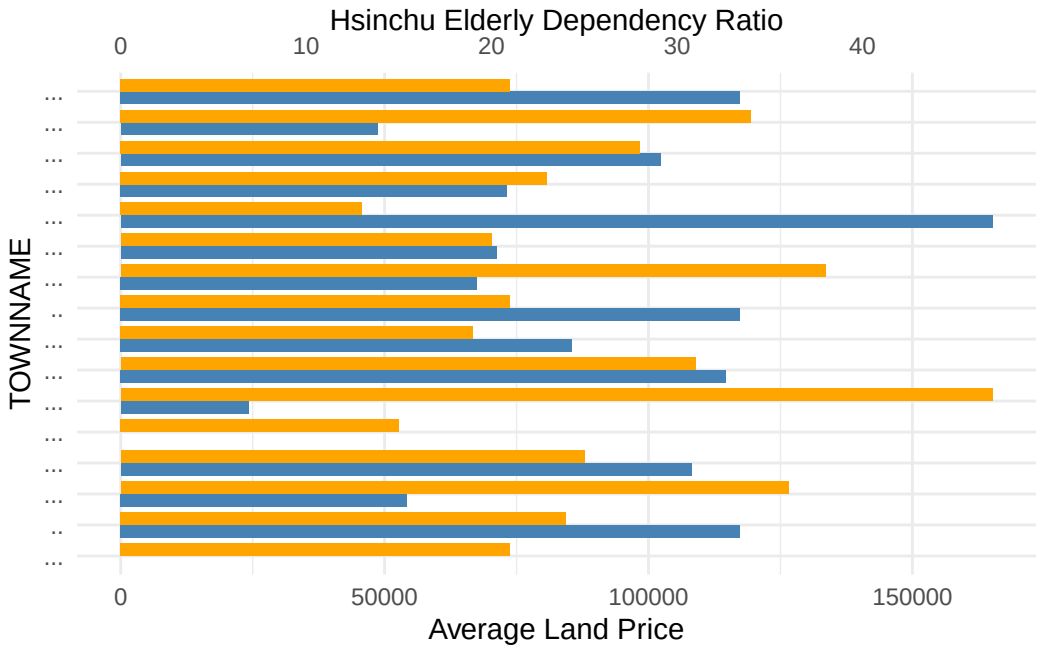
Bar Plot of Average Land Prices with Elderly Dependency Ratio as Second Scale

This plot demonstrates the comparison of the elderly dependency ratio vs average land prices per townships and districts in Hsinchu. This gives us an illustration of their relationship.

```
# Compute scaling factor
scale_factor <- max(hsinchu_map_data$avg_price, na.rm = TRUE) / max(hsinchu_map_data$A65UP_A

ggplot(hsinchu_map_data, aes(x = TOWNNAME)) +
  geom_col(aes(y = avg_price), fill = "steelblue", width = 0.4, position = position_nudge(x =
  geom_col(aes(y = A65UP_A15A64_RAT * scale_factor), fill = "orange", width = 0.4, position =
  scale_y_continuous(
    name = "Average Land Price",
    sec.axis = sec_axis(~./scale_factor, name = "Hsinchu Elderly Dependency Ratio")
  ) +
  coord_flip() +
  theme_minimal()
```

```
Warning: Removed 2 rows containing missing values or values outside the scale range
(`geom_col()`).
```



Scatter Plot

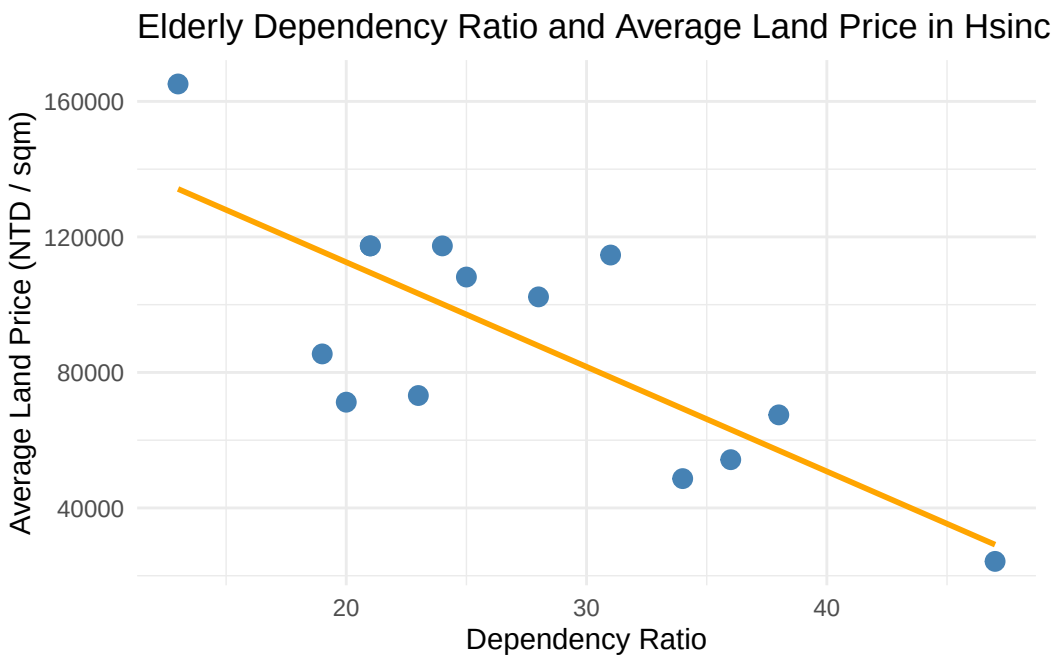
The scatter plot shows us a better picture of the trend between the relationship of the elderly dependency ratio and the average land prices. In this case, we can see that the increase of the average land prices as an inverse relationship to the elderly dependency ratio.

```
ggplot(hsinchu_map_data, aes(A65UP_A15A64_RAT, avg_price)) +
  geom_point(color = "steelblue", size = 3) +
  geom_smooth(method = "lm", color = "orange", se = FALSE) +
  labs(
    title = "Elderly Dependency Ratio and Average Land Price in Hsinchu",
    x = "Dependency Ratio",
    y = "Average Land Price (NTD / sqm)"
  ) +
  theme_minimal()
```

```
`geom_smooth()` using formula = 'y ~ x'
```

```
Warning: Removed 2 rows containing non-finite outside the scale range
(`stat_smooth()`).
```

```
Warning: Removed 2 rows containing missing values or values outside the scale range
(`geom_point()`).
```



Taipei City

We will do the same process to Taipei City

```
taipei_pop_data <- aging_data %>%
  filter(COUNTY %in% c("  ")) %>%
  select(2, 4, 10) %>%
  mutate(A65UP_A15A64_RAT = as.integer(A65UP_A15A64_RAT)) %>%
  rename(TOWNNAME = TOWN)

taipei_pop_data
```

```
# A tibble: 12 x 3
   COUNTY TOWNNAME A65UP_A15A64_RAT
   <chr>  <chr>                <int>
```

1	38
2	38
3	41
4	37
5	36
6	35
7	38
8	31
9	31
10	28
11	37
12	33

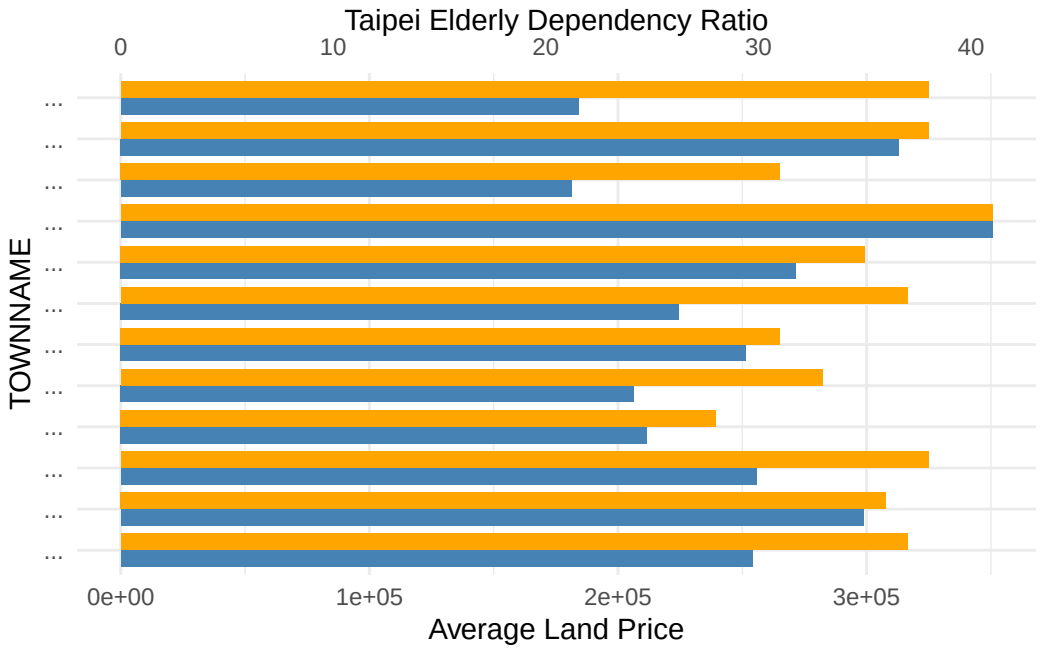
```
taipei_map_data <- taipei_map_data %>%
  left_join(taipei_pop_data, by = "TOWNNAME")


taipei_map_data$avg_price <- as.numeric(taipei_map_data$avg_price)
taipei_map_data$A65UP_A15A64_RAT <- as.numeric(taipei_map_data$A65UP_A15A64_RAT)
```

Bar Plot of Average Land Prices with Elderly Dependency Ratio as Second Scale

```
# Compute scaling factor
scale_factor <- max(taipei_map_data$avg_price, na.rm = TRUE) / max(taipei_map_data$A65UP_A15

ggplot(taipei_map_data, aes(x = TOWNNAME)) +
  geom_col(aes(y = avg_price), fill = "steelblue", width = 0.4, position = position_nudge(x =
  geom_col(aes(y = A65UP_A15A64_RAT * scale_factor), fill = "orange", width = 0.4, position =
  scale_y_continuous(
    name = "Average Land Price",
    sec.axis = sec_axis(~./scale_factor, name = "Taipei Elderly Dependency Ratio")
  ) +
  coord_flip() +
  theme_minimal()
```

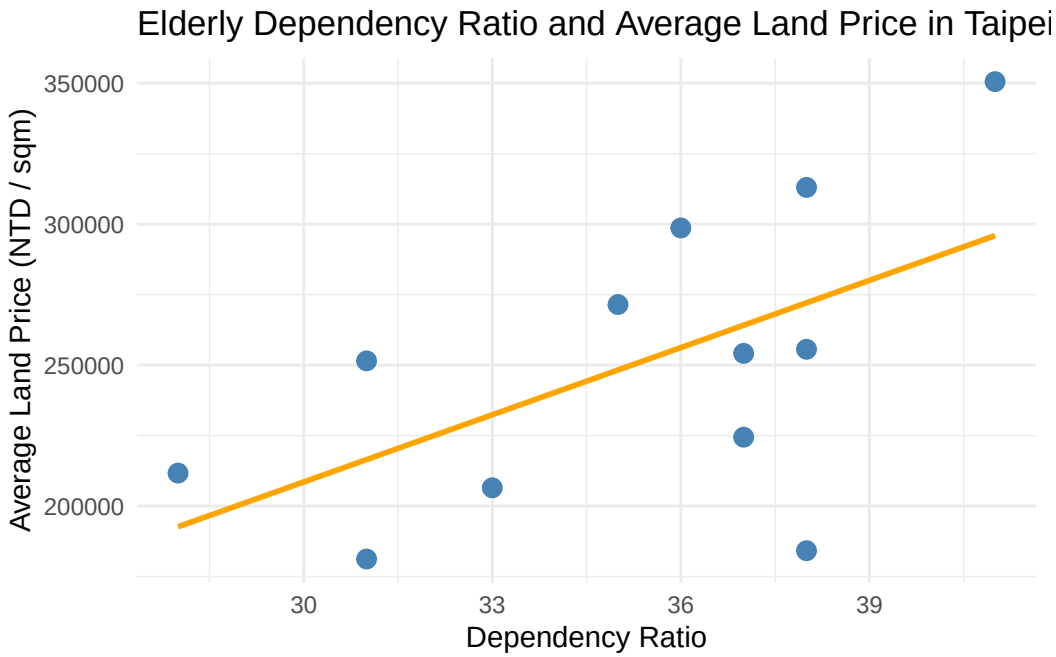


Scatter Plot

In this scatter ploy, we can see an interesting trand where the increase of land prices has a directly proportional to the elderly dependency ratio.

```
ggplot(taipei_map_data, aes(A65UP_A15A64_RAT, avg_price)) +
  geom_point(color = "steelblue", size = 3) +
  geom_smooth(method = "lm", color = "orange", se = FALSE) +
  labs(
    title = "Elderly Dependency Ratio and Average Land Price in Taipei",
    x = "Dependency Ratio",
    y = "Average Land Price (NTD / sqm)"
  ) +
  theme_minimal()
```

```
`geom_smooth()` using formula = 'y ~ x'
```


Elderly Dependency Ratio and Average Land Price in Taipei
Average Land Price (NTD / sqm)
350000
300000
250000
200000
30
33
36
39
Dependency Ratio