lj_investigate

wangsy

2025-09-15

1 Analysis of ideas and method

1.1 collect Shanghai Lianjia housing data to form an R dataset.

1.2 using RStudio tools and R code libraries for data analysis

1.3 it requires the use of many R language libraries, including the library collection: tidyverse, dplyr, and so on.

1.4 using R code, output a scatter plot showing the relationship between the construction age and floor area of houses.

2 Analysis Process

2.1 load lj data

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ---------------------- tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.2
## v ggplot2   4.0.0     v tibble    3.3.0
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.1.0
## -- Conflicts ------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

```r
library(dplyr)
library(knitr)
library(kableExtra)
```

```
##
## 载入程序包：'kableExtra'
##
## The following object is masked from 'package:dplyr':
##
##     group_rows
```

```r
library(formattable)
load("../data/lj_sh_2019.RData")
```

```r
# group statistics and calculate the proportion
housing_stats <- lj %>%
  mutate(building_age = 2025 - building_year) %>%
  filter(building_age <= 20) %>%
  group_by(building_age, building_area) %>%
  summarise(building_amount = n(), .groups = "drop") %>%
  mutate(
    proportion = paste0(round(building_amount/sum(building_amount)*100, 2), "%"),
    building_year = 2025-building_age
  ) %>%
  arrange(desc(proportion),desc(building_amount)) %>%
  select(building_year, building_age, building_area, building_amount, proportion)

# output table
knitr::kable(
  head(housing_stats, 300),
  caption = "statistics on the distribution of newly built houses within 30 years (sorted by regio
  align = c("l", "c", "c", "r", "r")
)
```

表 1: statistics on the distribution of newly built houses within
30 years (sorted by regional area)

building_year	building_age	building_area	building_amount	proportion
2006	19	100.69	6	0.46%
2011	14	43.27	3	0.23%
2011	14	50.45	3	0.23%
2011	14	87.68	3	0.23%
2011	14	89.15	3	0.23%
2011	14	122.65	3	0.23%
2011	14	129.94	3	0.23%
2010	15	80.09	3	0.23%
2010	15	99.64	3	0.23%
2009	16	76.63	3	0.23%
2009	16	83.93	3	0.23%
2009	16	92.47	3	0.23%
2007	18	76.80	3	0.23%
2007	18	96.53	3	0.23%
2007	18	102.21	3	0.23%
2006	19	170.00	3	0.23%
2005	20	66.15	3	0.23%
2005	20	260.44	3	0.23%
2015	10	91.48	2	0.15%
2015	10	282.00	2	0.15%
2014	11	99.30	2	0.15%
2014	11	171.00	2	0.15%
2014	11	488.00	2	0.15%
2013	12	45.93	2	0.15%
2013	12	74.76	2	0.15%
2012	13	47.26	2	0.15%
2012	13	74.96	2	0.15%
2012	13	82.19	2	0.15%
2012	13	119.64	2	0.15%
2012	13	138.34	2	0.15%
2012	13	138.78	2	0.15%
2012	13	162.00	2	0.15%
2012	13	174.97	2	0.15%
2012	13	227.13	2	0.15%
2012	13	238.93	2	0.15%

building_year	building_age	building_area	building_amount	proportion
2012	13	403.91	2	0.15%
2011	14	47.00	2	0.15%
2011	14	65.25	2	0.15%
2011	14	68.44	2	0.15%
2011	14	69.32	2	0.15%
2011	14	69.59	2	0.15%
2011	14	72.14	2	0.15%
2011	14	76.11	2	0.15%
2011	14	77.99	2	0.15%
2011	14	80.45	2	0.15%
2011	14	88.46	2	0.15%
2011	14	90.50	2	0.15%
2011	14	91.38	2	0.15%
2011	14	96.00	2	0.15%
2011	14	97.11	2	0.15%
2011	14	112.16	2	0.15%
2011	14	205.00	2	0.15%
2010	15	42.19	2	0.15%
2010	15	55.18	2	0.15%
2010	15	70.77	2	0.15%
2010	15	74.00	2	0.15%
2010	15	79.45	2	0.15%
2010	15	80.76	2	0.15%
2010	15	88.86	2	0.15%
2010	15	90.95	2	0.15%
2010	15	94.43	2	0.15%
2010	15	97.00	2	0.15%
2010	15	143.00	2	0.15%
2009	16	53.50	2	0.15%
2009	16	64.26	2	0.15%
2009	16	77.12	2	0.15%
2009	16	91.26	2	0.15%
2009	16	121.34	2	0.15%
2008	17	36.54	2	0.15%
2008	17	73.00	2	0.15%
2008	17	84.52	2	0.15%
2008	17	104.00	2	0.15%
2008	17	108.84	2	0.15%

building_year	building_age	building_area	building_amount	proportion
2008	17	123.46	2	0.15%
2008	17	130.04	2	0.15%
2008	17	130.81	2	0.15%
2008	17	145.39	2	0.15%
2008	17	164.55	2	0.15%
2008	17	325.73	2	0.15%
2007	18	52.88	2	0.15%
2007	18	66.11	2	0.15%
2007	18	69.56	2	0.15%
2007	18	71.43	2	0.15%
2007	18	88.18	2	0.15%
2007	18	92.58	2	0.15%
2007	18	95.00	2	0.15%
2007	18	95.56	2	0.15%
2007	18	96.20	2	0.15%
2007	18	109.46	2	0.15%
2007	18	115.00	2	0.15%
2007	18	115.67	2	0.15%
2007	18	131.16	2	0.15%
2007	18	138.67	2	0.15%
2006	19	47.55	2	0.15%
2006	19	97.35	2	0.15%
2006	19	117.98	2	0.15%
2006	19	119.17	2	0.15%
2006	19	123.07	2	0.15%
2006	19	133.77	2	0.15%
2006	19	134.00	2	0.15%
2006	19	136.00	2	0.15%
2006	19	137.00	2	0.15%
2006	19	139.25	2	0.15%
2006	19	140.10	2	0.15%
2006	19	150.88	2	0.15%
2006	19	192.45	2	0.15%
2006	19	251.00	2	0.15%
2006	19	314.55	2	0.15%
2006	19	324.00	2	0.15%
2005	20	36.34	2	0.15%
2005	20	39.10	2	0.15%

building_year	building_age	building_area	building_amount	proportion
2005	20	42.18	2	0.15%
2005	20	54.09	2	0.15%
2005	20	54.96	2	0.15%
2005	20	56.69	2	0.15%
2005	20	57.20	2	0.15%
2005	20	61.97	2	0.15%
2005	20	65.82	2	0.15%
2005	20	66.16	2	0.15%
2005	20	69.31	2	0.15%
2005	20	70.76	2	0.15%
2005	20	88.22	2	0.15%
2005	20	88.54	2	0.15%
2005	20	91.99	2	0.15%
2005	20	92.00	2	0.15%
2005	20	98.56	2	0.15%
2005	20	107.45	2	0.15%
2005	20	108.49	2	0.15%
2005	20	109.00	2	0.15%
2005	20	114.48	2	0.15%
2005	20	131.23	2	0.15%
2005	20	133.73	2	0.15%
2005	20	143.39	2	0.15%
2005	20	143.83	2	0.15%
2005	20	184.69	2	0.15%
2005	20	212.65	2	0.15%
2017	8	88.68	1	0.08%
2017	8	88.77	1	0.08%
2017	8	98.00	1	0.08%
2016	9	28.00	1	0.08%
2016	9	62.33	1	0.08%
2016	9	86.74	1	0.08%
2016	9	91.24	1	0.08%
2016	9	93.59	1	0.08%
2016	9	104.54	1	0.08%
2016	9	114.75	1	0.08%
2016	9	117.07	1	0.08%
2016	9	140.65	1	0.08%
2015	10	30.96	1	0.08%

building_year	building_age	building_area	building_amount	proportion
2015	10	31.00	1	0.08%
2015	10	39.62	1	0.08%
2015	10	52.73	1	0.08%
2015	10	60.00	1	0.08%
2015	10	67.78	1	0.08%
2015	10	69.40	1	0.08%
2015	10	78.00	1	0.08%
2015	10	78.46	1	0.08%
2015	10	79.00	1	0.08%
2015	10	79.68	1	0.08%
2015	10	81.52	1	0.08%
2015	10	81.58	1	0.08%
2015	10	84.50	1	0.08%
2015	10	87.00	1	0.08%
2015	10	90.00	1	0.08%
2015	10	90.02	1	0.08%
2015	10	91.00	1	0.08%
2015	10	93.89	1	0.08%
2015	10	98.60	1	0.08%
2015	10	101.00	1	0.08%
2015	10	120.00	1	0.08%
2015	10	124.35	1	0.08%
2015	10	129.22	1	0.08%
2015	10	144.41	1	0.08%
2015	10	147.16	1	0.08%
2015	10	148.21	1	0.08%
2015	10	167.82	1	0.08%
2015	10	167.88	1	0.08%
2015	10	170.00	1	0.08%
2015	10	184.53	1	0.08%
2015	10	188.00	1	0.08%
2015	10	293.86	1	0.08%
2015	10	369.00	1	0.08%
2015	10	423.34	1	0.08%
2014	11	37.93	1	0.08%
2014	11	43.86	1	0.08%
2014	11	44.00	1	0.08%
2014	11	48.05	1	0.08%

building_year	building_age	building_area	building_amount	proportion
2014	11	48.11	1	0.08%
2014	11	49.06	1	0.08%
2014	11	49.99	1	0.08%
2014	11	55.00	1	0.08%
2014	11	58.18	1	0.08%
2014	11	61.00	1	0.08%
2014	11	62.12	1	0.08%
2014	11	66.63	1	0.08%
2014	11	69.42	1	0.08%
2014	11	71.20	1	0.08%
2014	11	76.00	1	0.08%
2014	11	86.33	1	0.08%
2014	11	86.67	1	0.08%
2014	11	87.00	1	0.08%
2014	11	88.28	1	0.08%
2014	11	88.41	1	0.08%
2014	11	89.80	1	0.08%
2014	11	90.00	1	0.08%
2014	11	90.65	1	0.08%
2014	11	90.75	1	0.08%
2014	11	93.98	1	0.08%
2014	11	94.35	1	0.08%
2014	11	94.88	1	0.08%
2014	11	95.34	1	0.08%
2014	11	98.62	1	0.08%
2014	11	100.47	1	0.08%
2014	11	100.79	1	0.08%
2014	11	101.00	1	0.08%
2014	11	102.97	1	0.08%
2014	11	106.33	1	0.08%
2014	11	106.75	1	0.08%
2014	11	110.96	1	0.08%
2014	11	113.51	1	0.08%
2014	11	113.98	1	0.08%
2014	11	114.65	1	0.08%
2014	11	119.29	1	0.08%
2014	11	120.35	1	0.08%
2014	11	120.70	1	0.08%

building_year	building_age	building_area	building_amount	proportion
2014	11	126.00	1	0.08%
2014	11	129.30	1	0.08%
2014	11	132.78	1	0.08%
2014	11	136.65	1	0.08%
2014	11	139.98	1	0.08%
2014	11	141.12	1	0.08%
2014	11	143.19	1	0.08%
2014	11	144.00	1	0.08%
2014	11	144.44	1	0.08%
2014	11	146.68	1	0.08%
2014	11	166.78	1	0.08%
2014	11	171.11	1	0.08%
2014	11	172.40	1	0.08%
2014	11	176.36	1	0.08%
2014	11	183.82	1	0.08%
2014	11	186.00	1	0.08%
2014	11	187.00	1	0.08%
2014	11	187.83	1	0.08%
2014	11	191.20	1	0.08%
2014	11	197.71	1	0.08%
2014	11	217.39	1	0.08%
2014	11	240.10	1	0.08%
2014	11	244.44	1	0.08%
2014	11	322.00	1	0.08%
2014	11	332.00	1	0.08%
2013	12	42.99	1	0.08%
2013	12	43.63	1	0.08%
2013	12	44.00	1	0.08%
2013	12	45.00	1	0.08%
2013	12	45.07	1	0.08%
2013	12	45.96	1	0.08%
2013	12	51.10	1	0.08%
2013	12	53.23	1	0.08%
2013	12	53.25	1	0.08%
2013	12	54.63	1	0.08%
2013	12	55.65	1	0.08%
2013	12	55.92	1	0.08%
2013	12	56.76	1	0.08%

building_year	building_age	building_area	building_amount	proportion
2013	12	57.87	1	0.08%
2013	12	59.19	1	0.08%
2013	12	60.00	1	0.08%
2013	12	60.09	1	0.08%
2013	12	61.43	1	0.08%
2013	12	63.00	1	0.08%
2013	12	64.66	1	0.08%
2013	12	66.05	1	0.08%
2013	12	68.63	1	0.08%
2013	12	69.24	1	0.08%
2013	12	71.00	1	0.08%
2013	12	73.50	1	0.08%
2013	12	73.71	1	0.08%
2013	12	74.36	1	0.08%
2013	12	74.61	1	0.08%
2013	12	74.99	1	0.08%
2013	12	76.00	1	0.08%
2013	12	81.41	1	0.08%
2013	12	82.00	1	0.08%
2013	12	83.00	1	0.08%
2013	12	84.00	1	0.08%
2013	12	85.68	1	0.08%
2013	12	86.28	1	0.08%
2013	12	88.00	1	0.08%
2013	12	90.28	1	0.08%
2013	12	90.32	1	0.08%
2013	12	90.93	1	0.08%
2013	12	92.11	1	0.08%
2013	12	92.78	1	0.08%
2013	12	92.97	1	0.08%
2013	12	93.35	1	0.08%
2013	12	99.33	1	0.08%
2013	12	103.96	1	0.08%
2013	12	105.66	1	0.08%
2013	12	121.93	1	0.08%
2013	12	125.39	1	0.08%
2013	12	129.67	1	0.08%

2.2 plot lj geom_point chart