小薇向前冲

我的R语言学习之路

薇薇

NCU

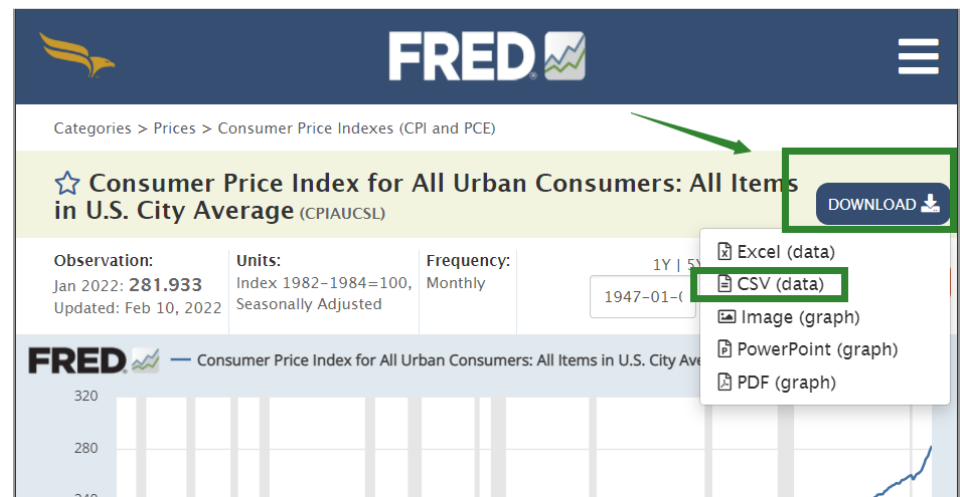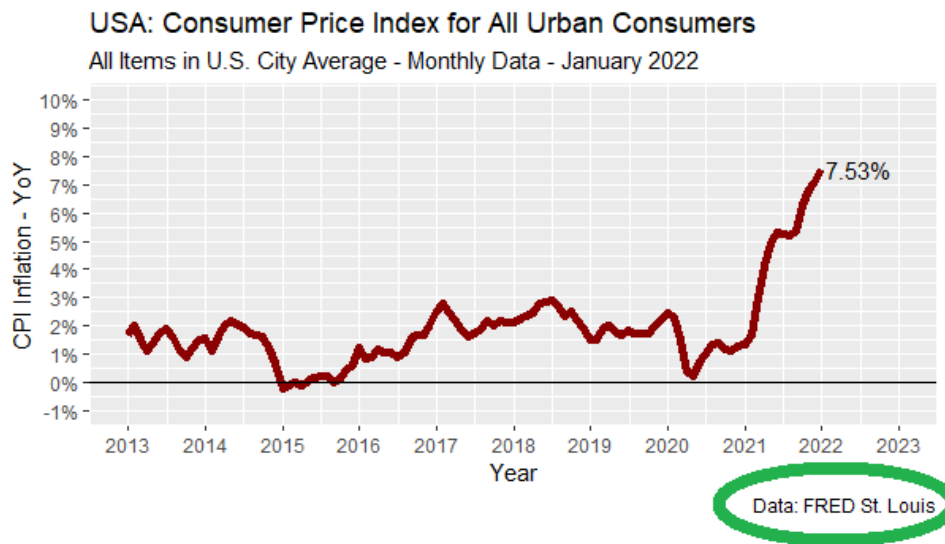2022-02-14

缘起

前几天看新闻的时候，发现一张图，勾起了我的兴趣，打算重复这张图
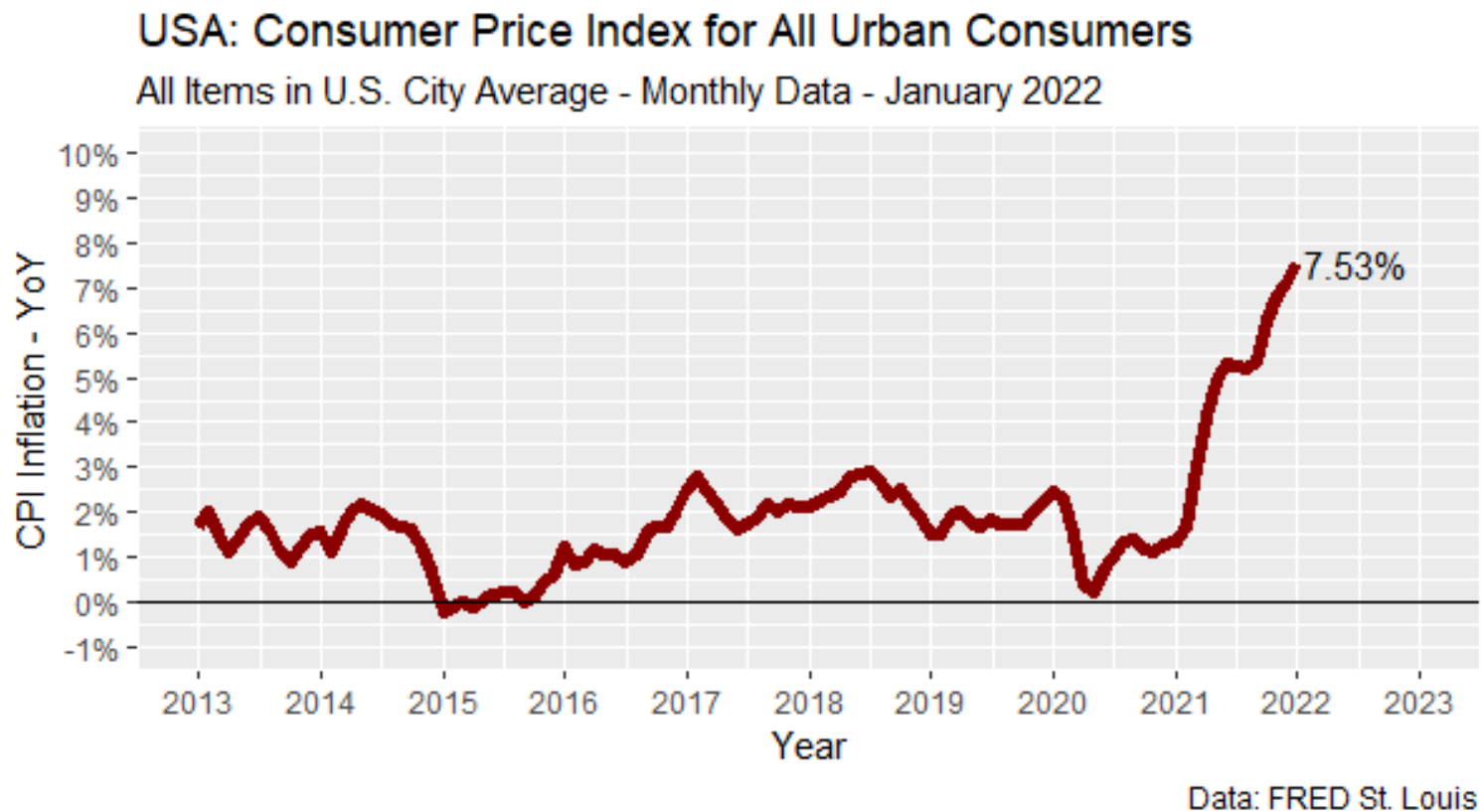
寻找数据

原图的右下角告诉了数据来源，于是找到这个网站https://fred.stlouisfed.org/series/CPIAUCSL，可获取数据

有了数据，就开工

读取csv文件用`readr::read_csv()` 很方便

```
library(tidyverse)
raw_df <- read_csv("./data/CPIAUCSL.csv")

raw_df
```

```
# A tibble: 145 x 2
   DATE        CPIAUCSL
   <date>         <dbl>
 1 2010-01-01     217.
 2 2010-02-01     217.
 3 2010-03-01     217.
 4 2010-04-01     217.
 5 2010-05-01     217.
 6 2010-06-01     217.
 7 2010-07-01     218.
 8 2010-08-01     218.
 9 2010-09-01     218.
10 2010-10-01     219.
# ... with 135 more rows
```

有了数据，就开工

读取csv文件用`readr::read_csv()` 很方便

```r
library(tidyverse)
raw_df <- read_csv("./data/CPIAUCSL.csv")

raw_df
```

```
# A tibble: 145 x 2
   DATE        CPIAUCSL
   <date>         <dbl>
 1 2010-01-01      217.
 2 2010-02-01      217.
 3 2010-03-01      217.
 4 2010-04-01      217.
 5 2010-05-01      217.
 6 2010-06-01      217.
 7 2010-07-01      218.
 8 2010-08-01      218.
 9 2010-09-01      218.
10 2010-10-01      219.
# ... with 135 more rows
```

列名推荐使用小写字母，阅读起来更舒服

```r
df <- raw_df %>%
  rename(date = DATE, cpi = CPIAUCSL)

df
```

```
# A tibble: 145 x 2
   date         cpi
   <date>      <dbl>
 1 2010-01-01  217.
 2 2010-02-01  217.
 3 2010-03-01  217.
 4 2010-04-01  217.
 5 2010-05-01  217.
 6 2010-06-01  217.
 7 2010-07-01  218.
 8 2010-08-01  218.
 9 2010-09-01  218.
10 2010-10-01  219.
# ... with 135 more rows
```

计算通货膨胀率

- 通货膨胀率计算公式

$$\text{Inflation (\%)} = \left(\frac{\text{Current CPI} - \text{Initial CPI}}{\text{Initial CPI}} \right) \times 100$$

$$\text{CPI} \longrightarrow \text{consumer price index}$$

- 注意 YoY 与 MoM 的区别
  - MoM: 这个月要和上个月比较
  - YoY: 今年的一月份要和去年的一月份比较，今年的二月份要和去年的二月份比较，...

计算通货膨胀率

这里使用了`dplyr::lag()`函数，向量往后退12格，目的是去年的一月份要和今年的一月份**对齐**

```
df %>%
  head(15)
```

```
# A tibble: 15 x 2
   date         cpi
   <date>     <dbl>
 1 2010-01-01  217.
 2 2010-02-01  217.
 3 2010-03-01  217.
 4 2010-04-01  217.
 5 2010-05-01  217.
 6 2010-06-01  217.
 7 2010-07-01  218.
 8 2010-08-01  218.
 9 2010-09-01  218.
10 2010-10-01  219.
11 2010-11-01  220.
12 2010-12-01  220.
13 2011-01-01  221.
14 2011-02-01  222.
15 2011-03-01  223.
```

```
df %>%
  mutate( cpi_base = lag(cpi, 12) )
```

```
# A tibble: 15 x 3
   date         cpi cpi_base
   <date>     <dbl>    <dbl>
 1 2010-01-01  217.       NA
 2 2010-02-01  217.       NA
 3 2010-03-01  217.       NA
 4 2010-04-01  217.       NA
 5 2010-05-01  217.       NA
 6 2010-06-01  217.       NA
 7 2010-07-01  218.       NA
 8 2010-08-01  218.       NA
 9 2010-09-01  218.       NA
10 2010-10-01  219.       NA
11 2010-11-01  220.       NA
12 2010-12-01  220.       NA
13 2011-01-01  221.      217.
14 2011-02-01  222.      217.
15 2011-03-01  223.      217.
```

计算通货膨胀率

然后按照公式计算通货膨胀率

```
tbl <- df %>%
  mutate(
    cpi_base = lag(cpi, 12),
    inflation_rate = (cpi - cpi_base) / cpi_base
  )
```

```
# A tibble: 6 x 4
  date          cpi cpi_base inflation_rate
  <date>      <dbl>    <dbl>          <dbl>
1 2010-01-01  217.      NA            NA
2 2010-02-01  217.      NA            NA
3 2010-03-01  217.      NA            NA
4 2011-01-01  221.     217.           0.0170
5 2011-02-01  222.     217.           0.0212
6 2011-03-01  223.     217.           0.0262
```

ggplot2 上场

```
tbl
```

```
# A tibble: 145 x 4
   date          cpi cpi_base inflation_rate
   <date>      <dbl>    <dbl>          <dbl>
 1 2010-01-01  217.       NA             NA
 2 2010-02-01  217.       NA             NA
 3 2010-03-01  217.       NA             NA
 4 2010-04-01  217.       NA             NA
 5 2010-05-01  217.       NA             NA
 6 2010-06-01  217.       NA             NA
 7 2010-07-01  218.       NA             NA
 8 2010-08-01  218.       NA             NA
 9 2010-09-01  218.       NA             NA
10 2010-10-01  219.       NA             NA
# ... with 135 more rows
```

```
tbl %>%
  filter(date >= "2013-01-01")
```

```
# A tibble: 109 x 4
   date          cpi cpi_base inflation_rate
   <date>      <dbl>    <dbl>          <dbl>
 1 2013-01-01  232.     228.         0.0168
 2 2013-02-01  233.     228.         0.0202
 3 2013-03-01  232.     229.         0.0152
 4 2013-04-01  232.     229.         0.0114
 5 2013-05-01  232.     229.         0.0139
 6 2013-06-01  232.     229.         0.0172
 7 2013-07-01  233.     229.         0.0189
 8 2013-08-01  233.     230.         0.0154
 9 2013-09-01  234.     231.         0.0109
10 2013-10-01  234.     232.         0.00877
# ... with 99 more rows
```
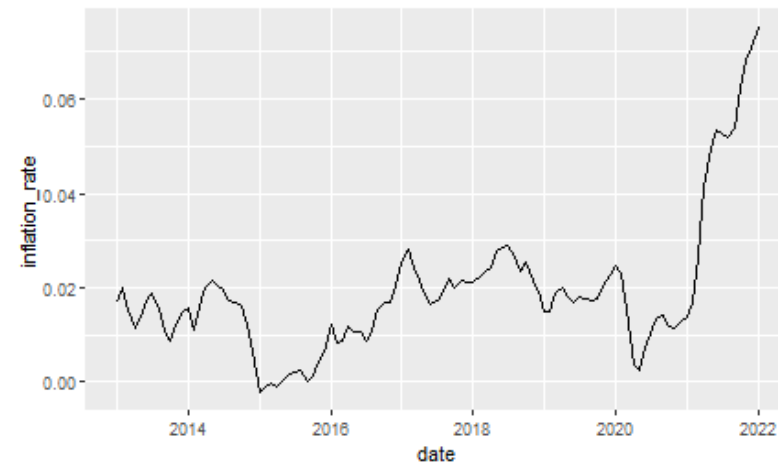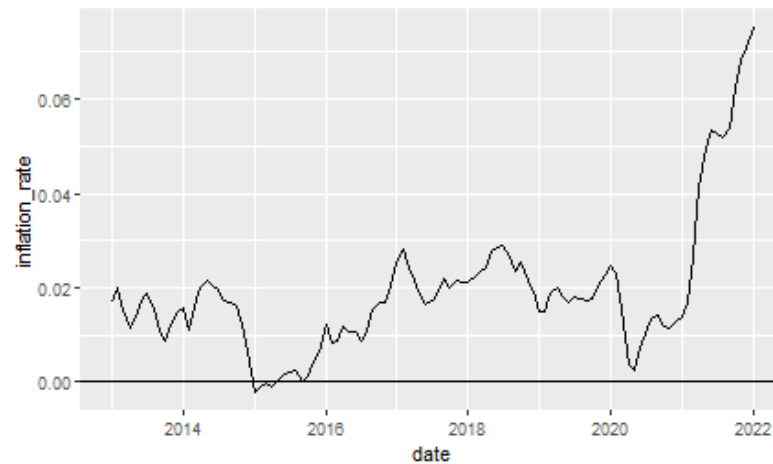
```
tbl %>%
  filter(date >= "2013-01-01") %>%
  ggplot(aes(x = date, y = inflation_rate))
```
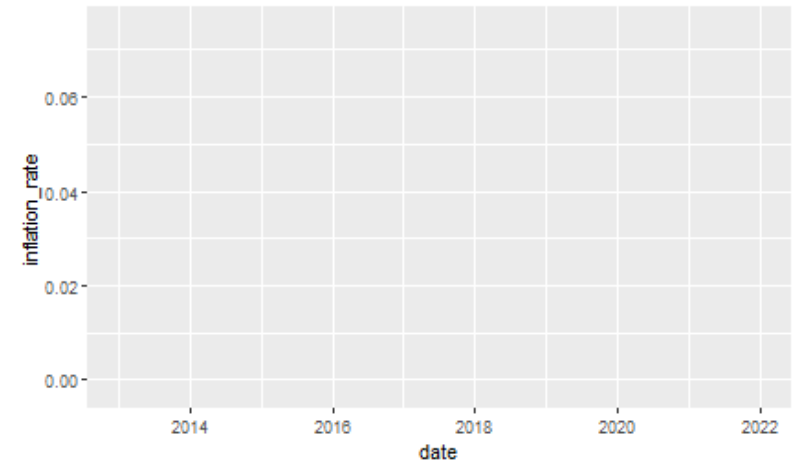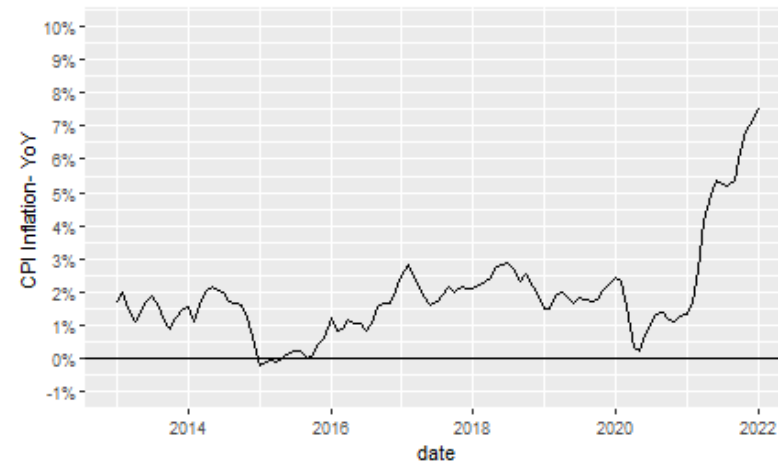
```
tbl %>%
  filter(date >= "2013-01-01") %>%
  ggplot(aes(x = date, y = inflation_rate)) +
  geom_line()
```

```
tbl %>%
  filter(date >= "2013-01-01") %>%
  ggplot(aes(x = date, y = inflation_rate)) +
  geom_line() +
  geom_hline(yintercept = 0)
```
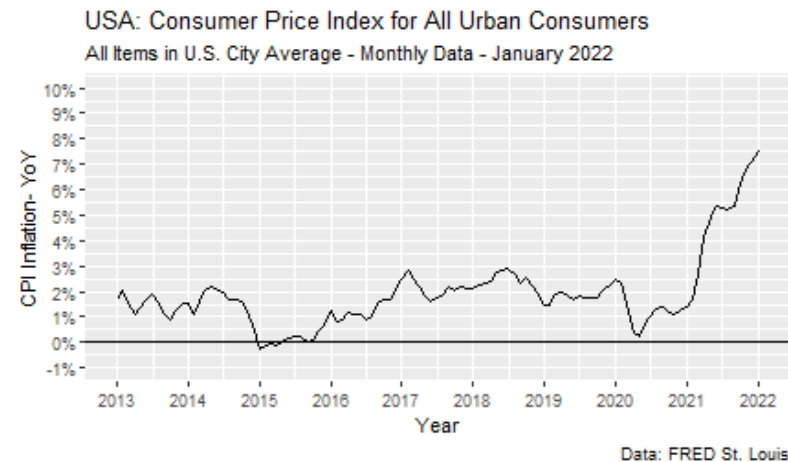
```
tbl %>%
  filter(date >= "2013-01-01") %>%
  ggplot(aes(x = date, y = inflation_rate)) +
  geom_line() +
  geom_hline(yintercept = 0) +
  scale_y_continuous(
    name = "CPI Inflation- YoY",
    limits = c(-0.01, 0.1),
    breaks = seq(from = -0.01, to = 0.1, by = 0.01),
    labels = scales::label_percent(scale = 100, accuracy
  )
```

```
tbl %>%
  filter(date >= "2013-01-01") %>%
  ggplot(aes(x = date, y = inflation_rate)) +
  geom_line() +
  geom_hline(yintercept = 0) +
  scale_y_continuous(
    name = "CPI Inflation- YoY",
    limits = c(-0.01, 0.1),
    breaks = seq(from = -0.01, to = 0.1, by = 0.01),
    labels = scales::label_percent(scale = 100, accuracy
  ) +
  scale_x_date(
    name = "Year",
    date_breaks = "1 year",
    date_labels = "%Y"
  )
```
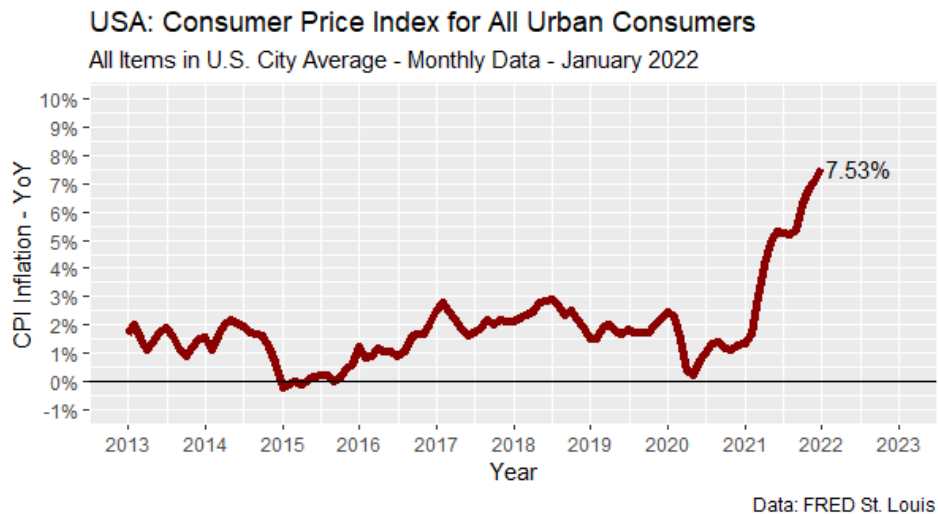
```r
tbl %>%
  filter(date >= "2013-01-01") %>%
  ggplot(aes(x = date, y = inflation_rate)) +
  geom_line() +
  geom_hline(yintercept = 0) +
  scale_y_continuous(
    name = "CPI Inflation- YoY",
    limits = c(-0.01, 0.1),
    breaks = seq(from = -0.01, to = 0.1, by = 0.01),
    labels = scales::label_percent(scale = 100, accuracy
  ) +
  scale_x_date(
    name = "Year",
    date_breaks = "1 year",
    date_labels = "%Y"
  ) +
  labs(
    title = "USA: Consumer Price Index for All Urban Cons
    subtitle = "All Items in U.S. City Average - Monthly
    caption = "Data: FRED St. Louis"
  )
```


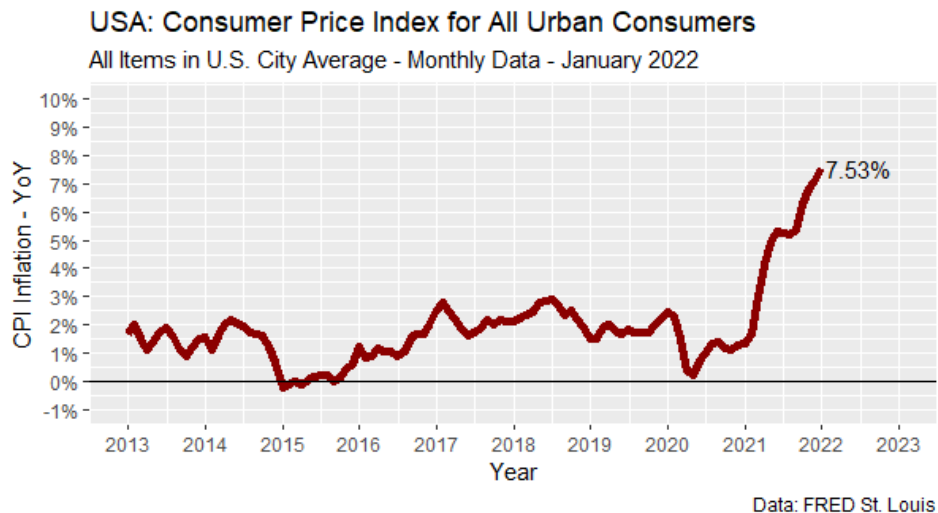USA: Consumer Price Index for All Urban Consumers
All Items in U.S. City Average - Monthly Data - January 2022

Data: FRED St. Louis

细节是魔鬼

原图的线条是什么颜色?

- 原图

原图的线条是什么颜色？

- 原图

- 哪个是你大门的颜色？

取色器
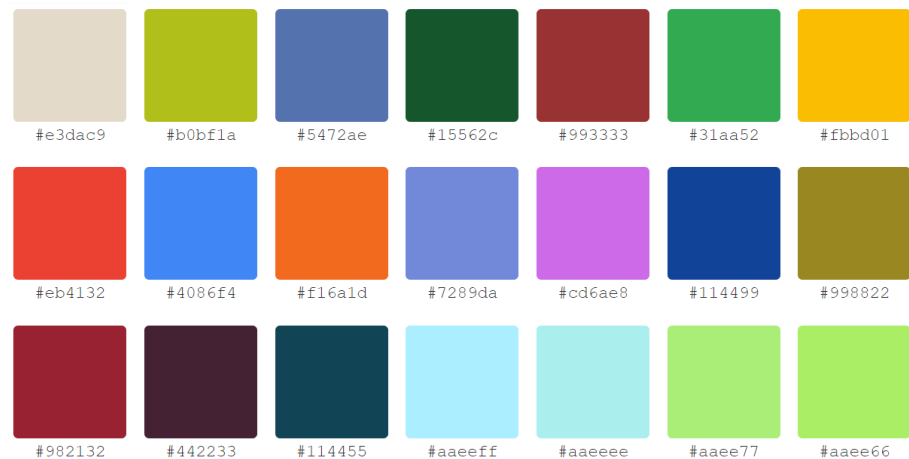
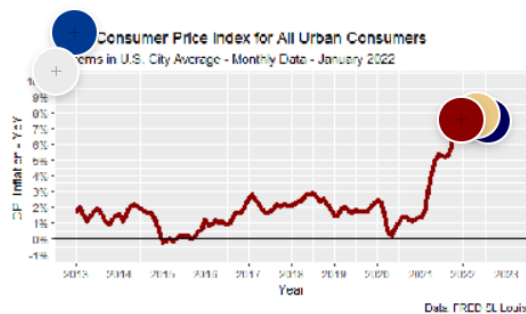- 取色器https://color.adobe.com/zh/create/color-wheel
- 把图片拉进去，可以得到图片中颜色hex值，是不是很神奇？

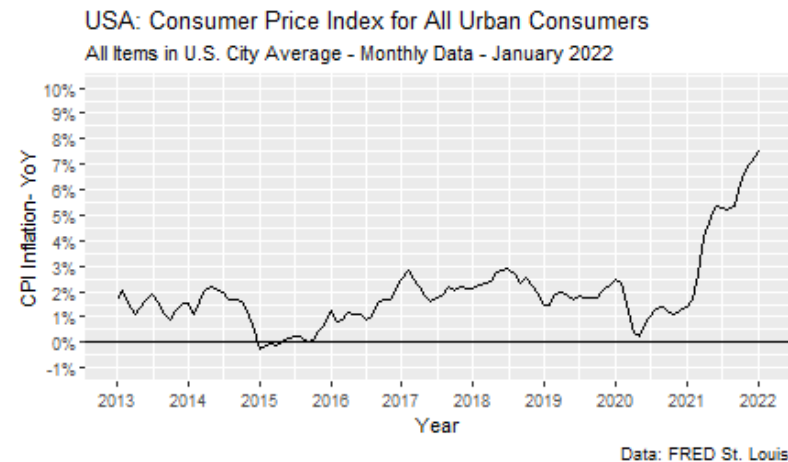#020659   #033E8C   #F2CF8D   #8C0303   #F2F2F2
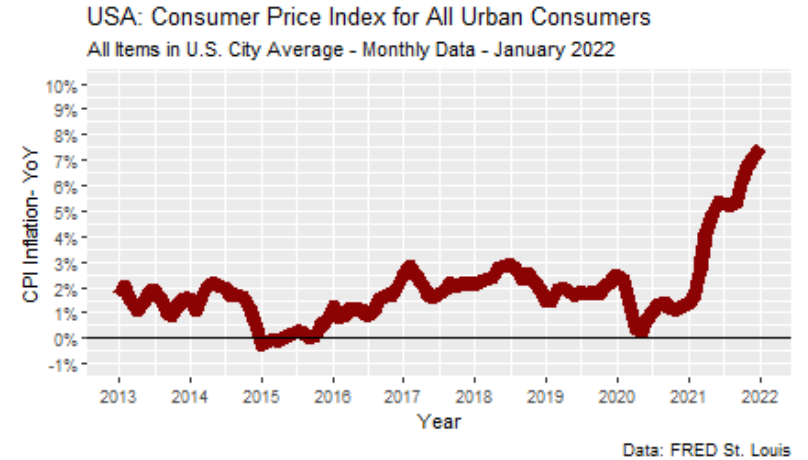
```
tbl %>%
  filter(date >= "2013-01-01") %>%
  ggplot(aes(x = date, y = inflation_rate)) +
  geom_line() +
  geom_hline(yintercept = 0) +
  scale_y_continuous(
    name = "CPI Inflation- YoY",
    limits = c(-0.01, 0.1),
    breaks = seq(from = -0.01, to = 0.1, by = 0.01),
    labels = scales::label_percent(scale = 100, accuracy
  ) +
scale_x_date(
    name = "Year",
    date_breaks = "1 year",
    date_labels = "%Y"
  ) +
labs(
    title = "USA: Consumer Price Index for All Urban Cons
    subtitle = "All Items in U.S. City Average - Monthly
    caption = "Data: FRED St. Louis"
  )
```


USA: Consumer Price Index for All Urban Consumers
All Items in U.S. City Average - Monthly Data - January 2022

Data: FRED St. Louis

```
tbl %>%
  filter(date >= "2013-01-01") %>%
  ggplot(aes(x = date, y = inflation_rate)) +
  geom_line(color = "#8C0303", size = 3) +
  geom_hline(yintercept = 0) +
  scale_y_continuous(
    name = "CPI Inflation- YoY",
    limits = c(-0.01, 0.1),
    breaks = seq(from = -0.01, to = 0.1, by = 0.01),
    labels = scales::label_percent(scale = 100, accuracy
  ) +
  scale_x_date(
    name = "Year",
    date_breaks = "1 year",
    date_labels = "%Y"
  ) +
  labs(
    title = "USA: Consumer Price Index for All Urban Cons
    subtitle = "All Items in U.S. City Average - Monthly
    caption = "Data: FRED St. Louis"
  )
```


USA: Consumer Price Index for All Urban Consumers
All Items in U.S. City Average - Monthly Data - January 2022

Data: FRED St. Louis
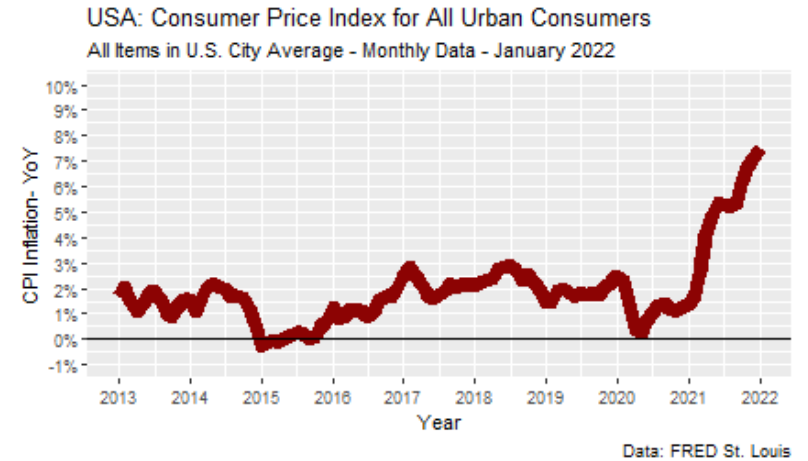
添加标签

将最后观察点的值弄成标签

方法：构建新的一列，如果是最后时间点，就放入该时间点的inflation_rate，否则，啥都不放。

```
tbl %>%
  mutate(
    endpoint = if_else(date == max(date), inflation_rate, NA_real_)
  ) %>%
  tail()
```

```
# A tibble: 6 x 5
  date          cpi cpi_base inflation_rate endpoint
  <date>      <dbl>    <dbl>          <dbl>    <dbl>
1 2021-08-01  273.     260.          0.0521       NA
2 2021-09-01  274.     260.          0.0539       NA
3 2021-10-01  277.     260.          0.0624       NA
4 2021-11-01  279.     261.          0.0683       NA
5 2021-12-01  280.     262.          0.0710       NA
6 2022-01-01  282.     262.          0.0753   0.0753
```
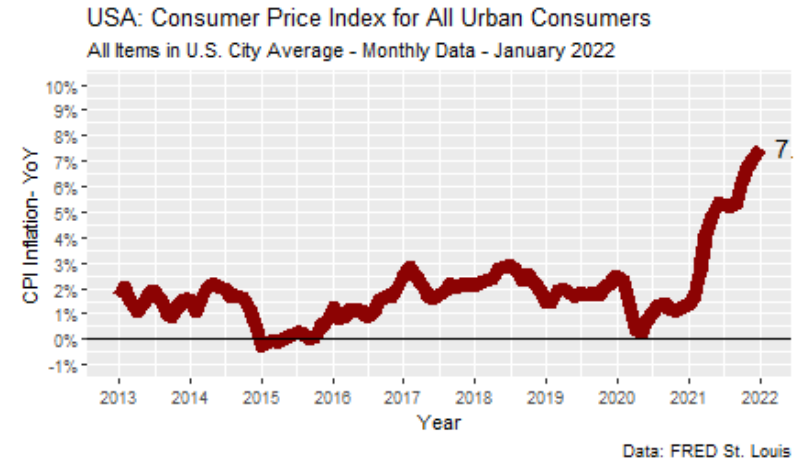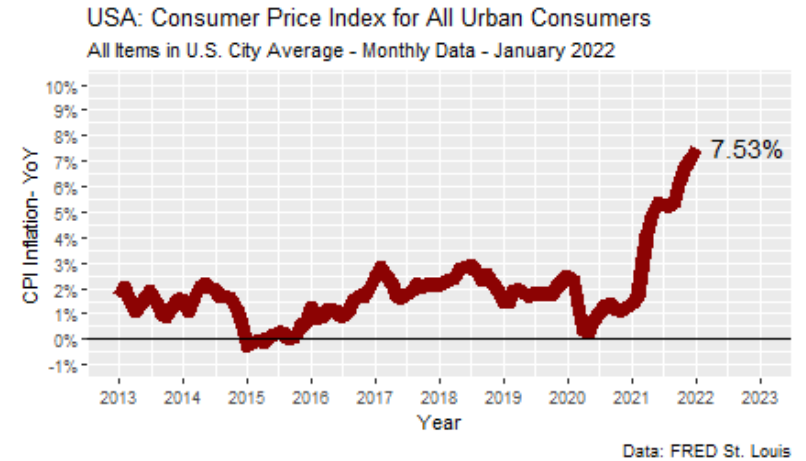
```
tbl %>%
  filter(date >= "2013-01-01") %>%
  ggplot(aes(x = date, y = inflation_rate)) +
  geom_line(color = "#8C0303", size = 3) +
  geom_hline(yintercept = 0) +
  scale_y_continuous(
    name = "CPI Inflation- YoY",
    limits = c(-0.01, 0.1),
    breaks = seq(from = -0.01, to = 0.1, by = 0.01),
    labels = scales::label_percent(scale = 100, accuracy
  ) +
  scale_x_date(
    name = "Year",
    date_breaks = "1 year",
    date_labels = "%Y"
  ) +
  labs(
    title = "USA: Consumer Price Index for All Urban Cons
    subtitle = "All Items in U.S. City Average - Monthly
    caption = "Data: FRED St. Louis"
  )
```


USA: Consumer Price Index for All Urban Consumers
All Items in U.S. City Average - Monthly Data - January 2022
Data: FRED St. Louis

```r
tbl %>%
  filter(date >= "2013-01-01") %>%
  mutate(endpoint = if_else(date == max(date), inflation_
  ggplot(aes(x = date, y = inflation_rate)) +
  geom_line(color = "#8C0303", size = 3) +
  geom_hline(yintercept = 0) +
  geom_text(
    aes(label = scales::percent(endpoint, accuracy = 0.0
    hjust = -0.2,
    size = 5
  ) +
  scale_y_continuous(
    name = "CPI Inflation- YoY",
    limits = c(-0.01, 0.1),
    breaks = seq(from = -0.01, to = 0.1, by = 0.01),
    labels = scales::label_percent(scale = 100, accuracy
  ) +
  scale_x_date(
    name = "Year",
    date_breaks = "1 year",
    date_labels = "%Y"
  ) +
  labs(
    title = "USA: Consumer Price Index for All Urban Cons
    subtitle = "All Items in U.S. City Average - Monthly
    caption = "Data: FRED St. Louis"
  )
```


USA: Consumer Price Index for All Urban Consumers
All Items in U.S. City Average - Monthly Data - January 2022
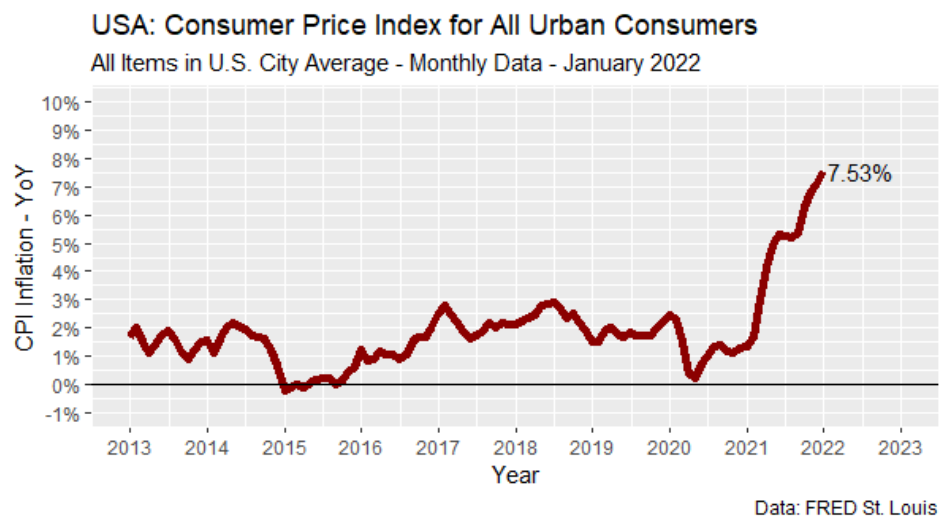
Data: FRED St. Louis

```
tbl %>%
  filter(date >= "2013-01-01") %>%
  mutate(endpoint = if_else(date == max(date), inflation
  ggplot(aes(x = date, y = inflation_rate)) +
  geom_line(color = "#8C0303", size = 3) +
  geom_hline(yintercept = 0) +
  geom_text(
    aes(label = scales::percent(endpoint, accuracy = 0.0
    hjust = -0.2,
    size = 5
  ) +
  scale_y_continuous(
    name = "CPI Inflation- YoY",
    limits = c(-0.01, 0.1),
    breaks = seq(from = -0.01, to = 0.1, by = 0.01),
    labels = scales::label_percent(scale = 100, accuracy
  ) +
  scale_x_date(
    name = "Year",
    limits = as.Date(c("2013-01-01", "2023-01-01")),
    date_breaks = "1 year",
    date_labels = "%Y"
  ) +
  labs(
    title = "USA: Consumer Price Index for All Urban Cons
    subtitle = "All Items in U.S. City Average - Monthly
    caption = "Data: FRED St. Louis"
  )
```
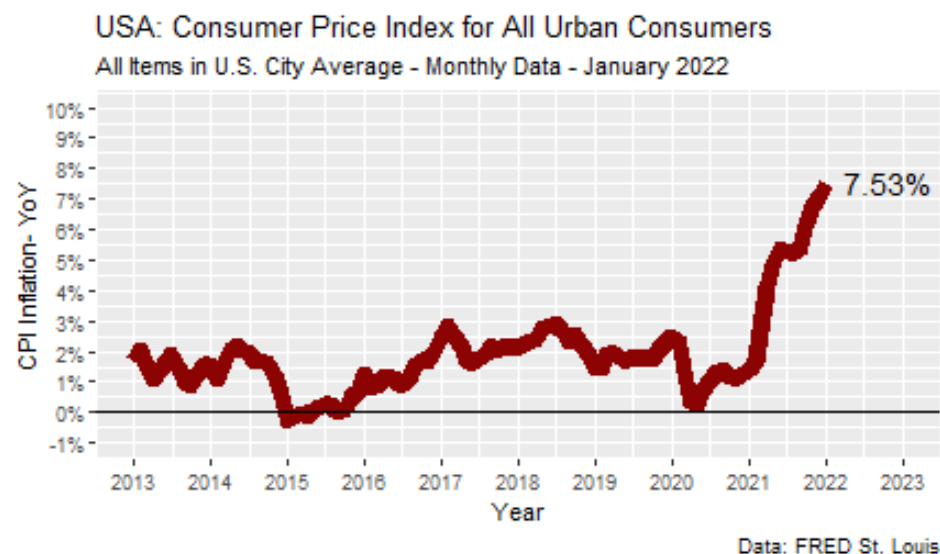

USA: Consumer Price Index for All Urban Consumers
All Items in U.S. City Average - Monthly Data - January 2022

7.53%

CPI Inflation- YoY

Year

Data: FRED St. Louis

安能辨我是雄雌

- 这是原图

- 这是重复的

参考

- https://fred.stlouisfed.org/series/CPIAUCSL
- https://www.bbc.com/zhongwen/simp/business-60345552
- https://www.econ.iastate.edu/ask-an-economist/cpi-and-inflation-relationship-between-mom-and-yoy-values
- https://ggplot2.tidyverse.org/

感谢 R 语言的美!

本幻灯片由 R 包 **xaringan** 和 **flipbookr** 生成