

Task: LOESS and GAM

Hongjin Lyu B00978648

2025-02-25

In this task you will practice fitting LOESS and GAM smooths and adding them to visualizations.

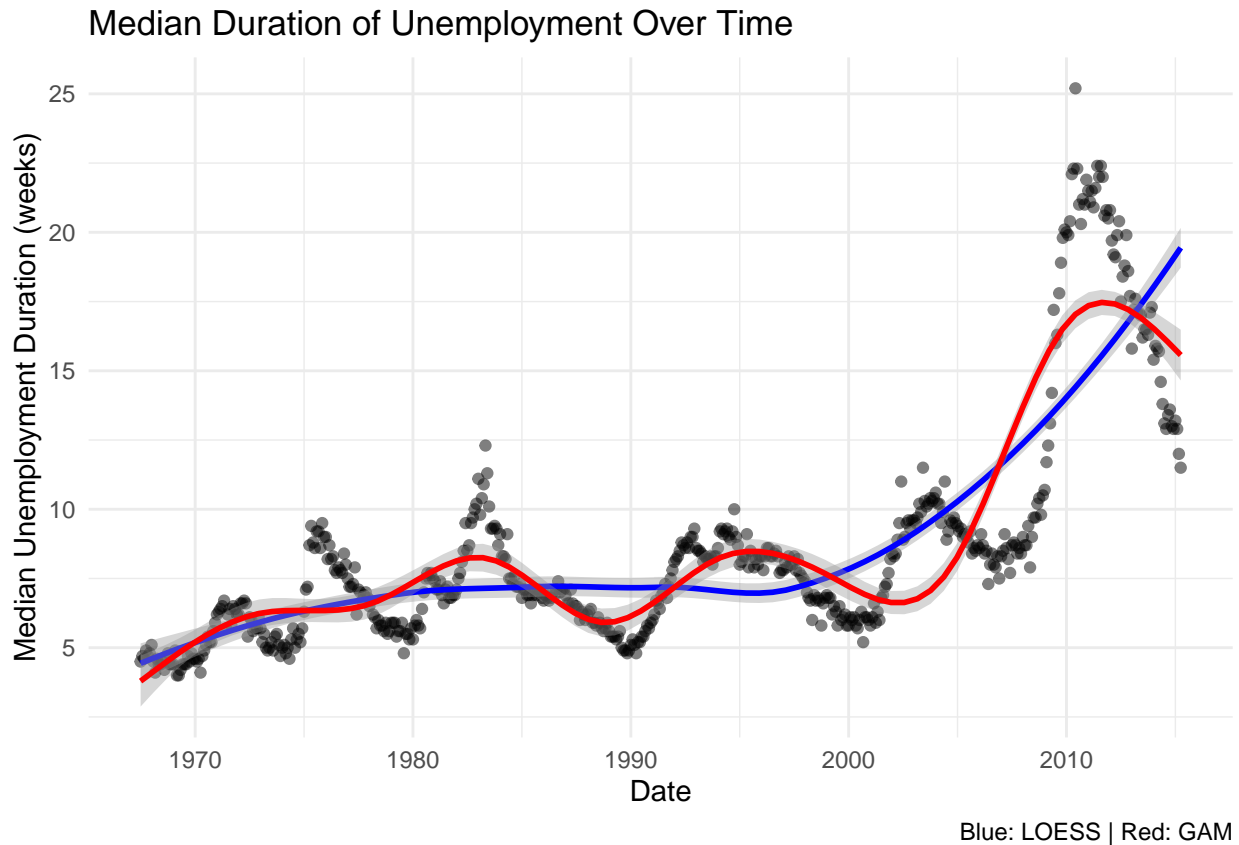
Question 1

Using the same data from the Linear Models Task, plot the median duration of unemployment as a function of the date. Add a GAM or LOESS smooth to the data which you think captures the main trend in the data.

```
data(economics)

library(ggplot2)
ggplot(economics, aes(x = date, y = uempmed)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "loess", color = "blue", se = TRUE) +
  geom_smooth(method = "gam", formula = y ~ s(x, bs = "cs"), color = "red", se = TRUE) +
  labs(title = "Median Duration of Unemployment Over Time",
       x = "Date",
       y = "Median Unemployment Duration (weeks)",
       caption = "Blue: LOESS | Red: GAM") +
  theme_minimal()

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



Explain your preferred choice. (Why is GAM better than LOESS, or the reverse?) What is the main trend highlighted by your smooth?

LOESS is useful for capturing short-term fluctuations and local trends in data. Since unemployment duration may exhibit nonlinear and short-term variations, LOESS can provide a flexible, non-parametric smoothing without assuming a specific functional form.

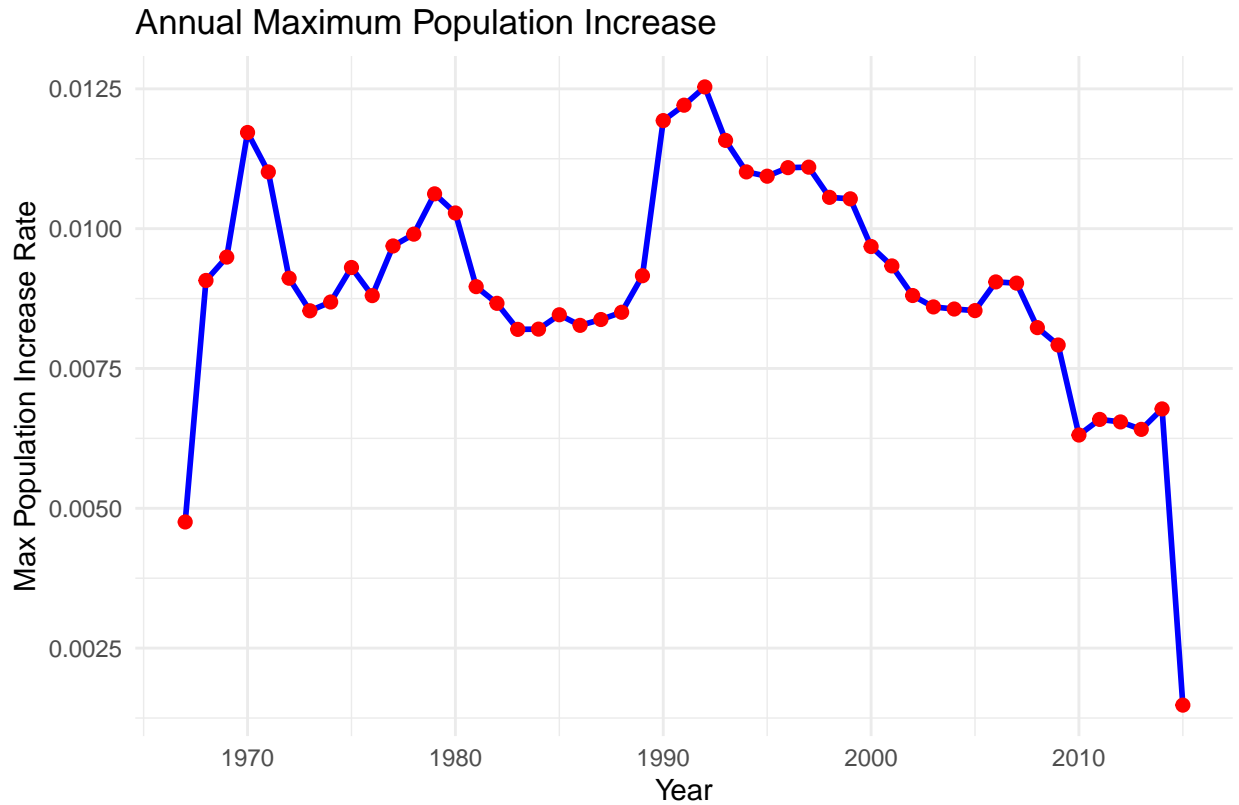
Question 2

In the code block below I calculate the increase in population each year, relative to lowest population in the year. The first and last year in the dataset are outliers because there is less than a full year of data for those years.

```
my_econ3 <- economics |>
  mutate(year = year(date), month = month(date)) |>
  group_by(year) |>
  mutate(pop_incr = (pop - min(pop)) / min(pop)) |>
  filter(pop_incr == max(pop_incr))

ggplot(my_econ3, aes(x = year, y = pop_incr)) +
  geom_line(color = "blue", size = 1) +
  geom_point(size = 2, color = "red") +
  labs(title = "Annual Maximum Population Increase",
       x = "Year",
       y = "Max Population Increase Rate",
       caption = "Data from `economics` dataset") +
  theme_minimal()
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



Data from 'economics' dataset

Plot the population increase (`pop_incr`) as a function of the year. Use `filter` to exclude the first and last years. Add a smooth curve to the graph. Choose the type you did not use in question 1 (e.g., if you used a GAM in question 1, use a LOESS here.)

```
my_econ3 <- economics |>
  mutate(year = year(date), month = month(date)) |>
  group_by(year) |>
  mutate(pop_incr = (pop - min(pop)) / min(pop)) |>
  filter(pop_incr == max(pop_incr))

print(range(my_econ3$year))

## [1] 1967 2015

filtered_econ <- my_econ3 |>
  filter(year > min(my_econ3$year) & year < max(my_econ3$year))

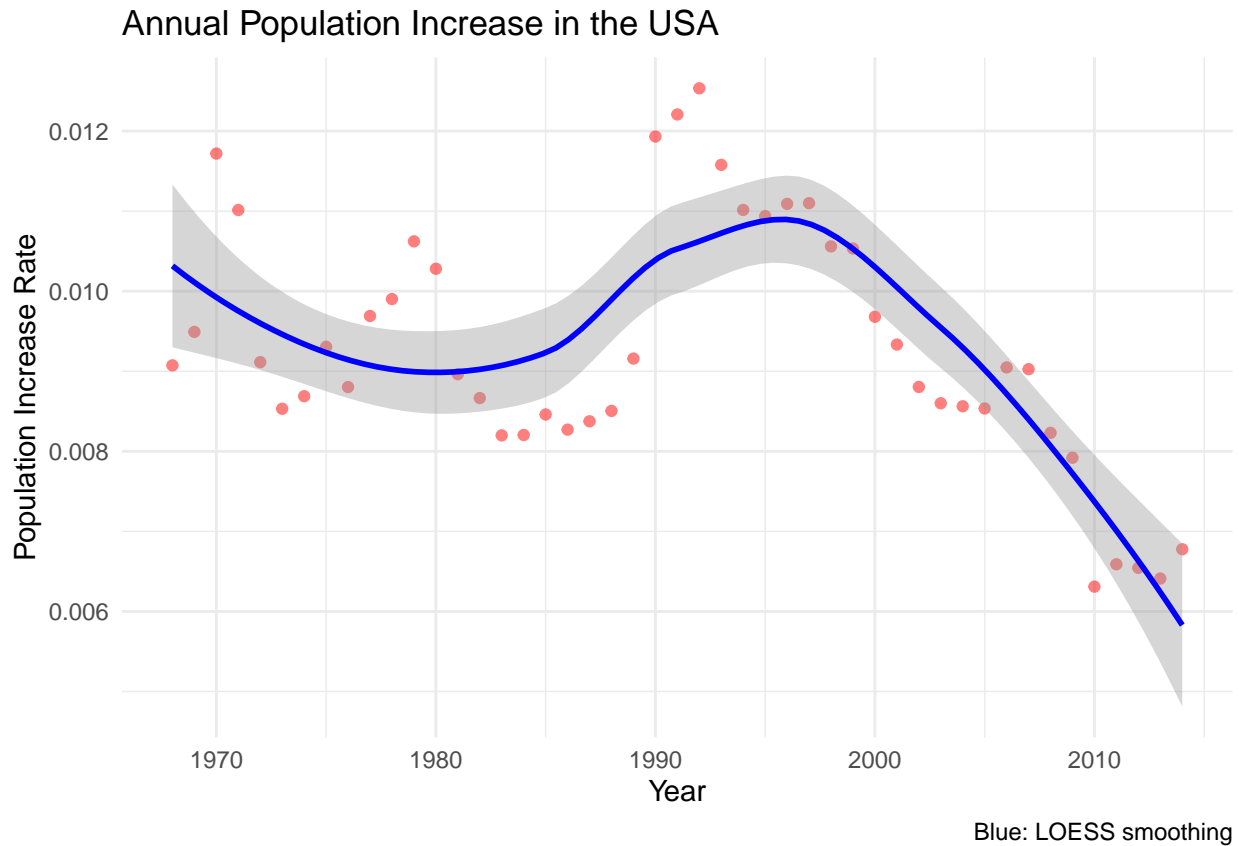
print(dim(filtered_econ))

## [1] 47 9

ggplot(filtered_econ, aes(x = year, y = pop_incr)) +
  geom_point(alpha = 0.5, color = "red") +
```

```
geom_smooth(method = "loess", color = "blue", se = TRUE) +
labs(title = "Annual Population Increase in the USA",
      x = "Year",
      y = "Population Increase Rate",
      caption = "Blue: LOESS smoothing") +
theme_minimal()
```

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



Summarize the trend in the annual rate of population increase in the USA in a sentence or two.

The annual rate of population increase in the USA shows a declining trend over time, with some fluctuations. This suggests that population growth is slowing down, possibly due to lower birth rates and other demographic factors.