Lab 11 - Data, Aesthetics, & Geometries

Ramin Jabbarialghanab

November 9, 2017

Complete the following exercises below. Knit together the PDF document and commit both the Lab 11 RMD file and the PDF document to Git. Push the changes to GitHub so both documents are visible in your public GitHub repository.

1. Which variables in your dataset are you interested in visualizing? Describe the level of measurement of these variables and what type of geography you think is appropriate to represent these variables. Give your reasoning for choosing the `geom_()` you selected.

I would visualize the variables of "year" and "value" in my data. The "value" varies based on "gender". On the X axis, I would have "year" and on the Y axis the values. The level of measurment for "year" is interval and the level of measurement for "value" is ratio. The gender(men and women) is nominal.

2. Is your data in the proper format to visualize the data in the way you want? Why or why not? *If you need/want to change the structure of your data, do it below.* Yes, my data is in the proper fomat.

3. Create at least two different exploratory plots of the variables you chose using the skills we covered in class today. What types of mapping aesthetics did you choose and why? What do these plots tell you about your data?

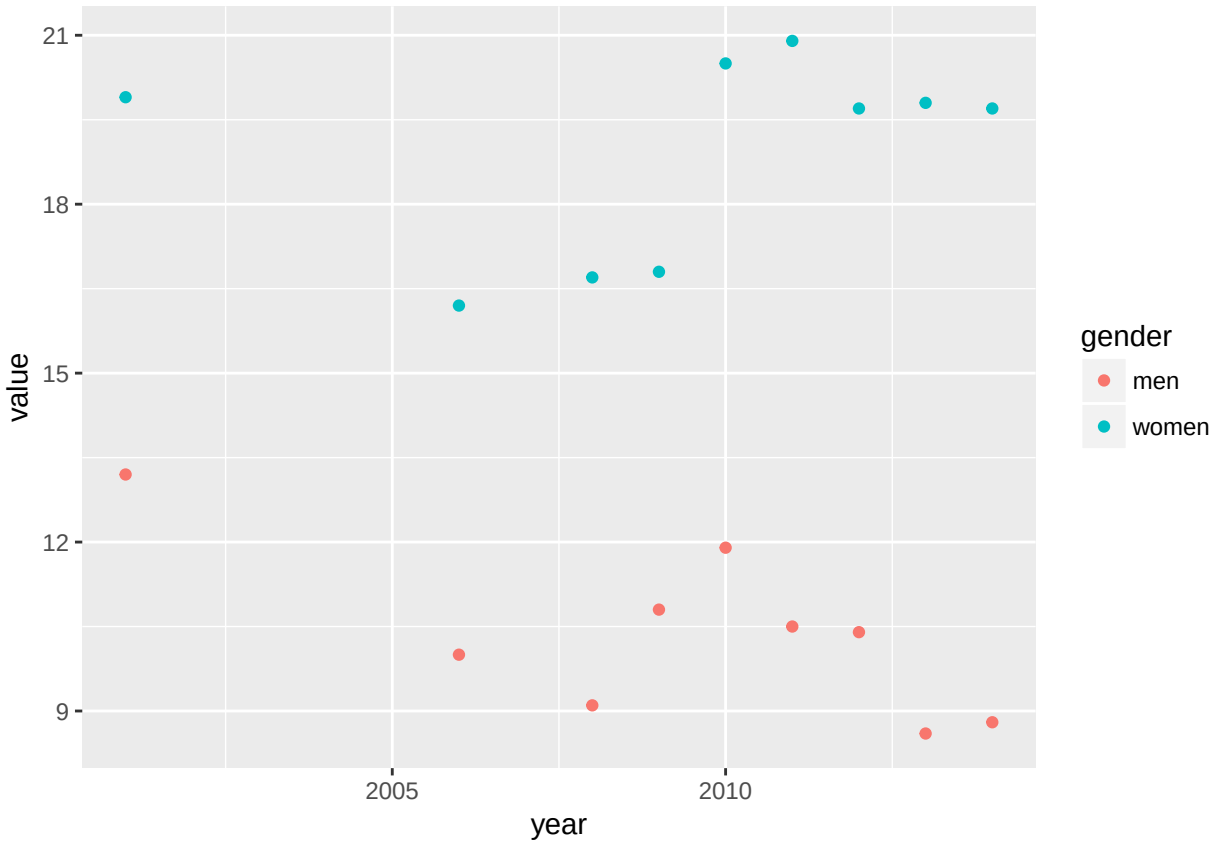
Since I need to compare "two" types of gender, I use x axis position for "year" variable and Y axis position for "value" variable. In addition, in order to distinguish men and women on the plot I uses "col" or "fill" for gender to have different colour for men and women. Here, for practice, I have more than two plots. The useful plots show that the rate of unemployment for women was alway more than the rate for women. Although since 2011 we have relatively sharp decreasing in the rate of unemployment for men, that does not happen for women. Overall chronologically the rate of women unemployment in comparison to the rate of men unemployment has increased. The difference is incresead.

```
# Creating geom point with col for gender
install.packages("ggplot2")
```

```
## Installing package into '/Users/Ramin/Desktop/Autumn 2017/Statistics 321/unemployment rate/unemploym
## (as 'lib' is unspecified)
```

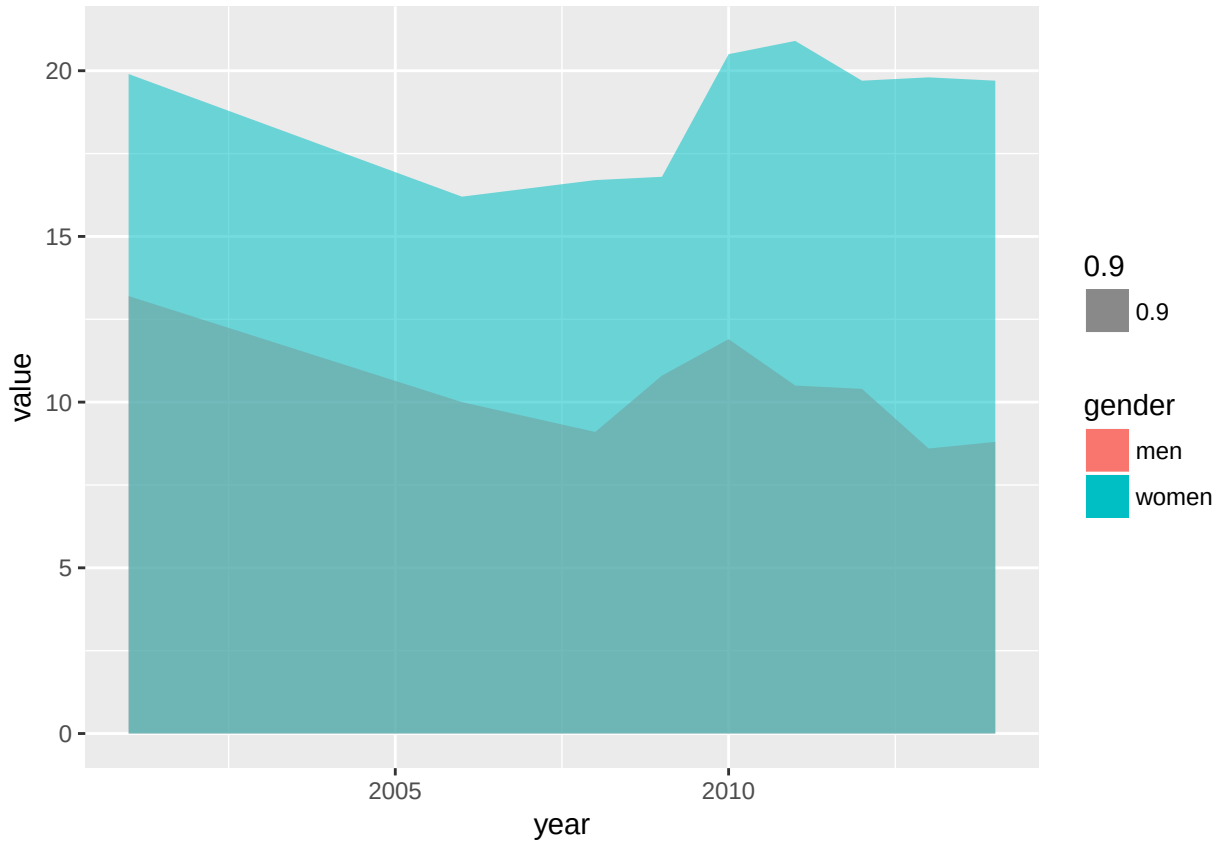
```
##
## The downloaded binary packages are in
##  /var/folders/z9/b9hh4hpj6hl6x9r9dxjctv240000gn/T//RtmpXmyjVp/downloaded_packages
```

```
library(ggplot2)
library(readxl)
joint_unemployment_total <- read_excel("~/Desktop/Autumn 2017/Statistics 321/unemployment rate/joint_un

ggplot(joint_unemployment_total,
       aes(x = year,
           y = value,
           col = gender)) +
  geom_point()
```

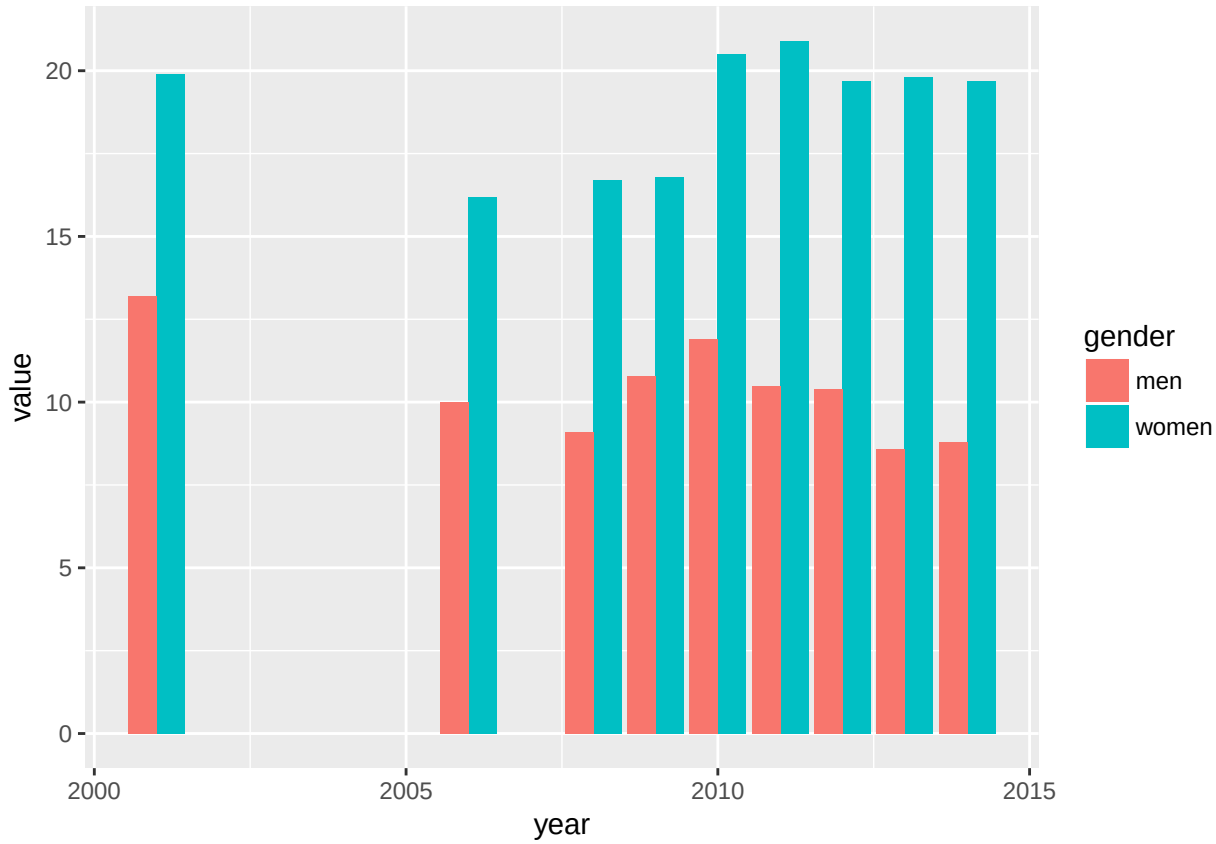



```
#creating geom ribbon with fill for gender

ggplot(joint_unemployment_total,
       aes(x = year,
           y = value,
          fill = gender)) +
  geom_ribbon(aes(ymax = value, xmax = year,
                  ymin = 0,
                  alpha = 0.9),
              position = "dodge")
```

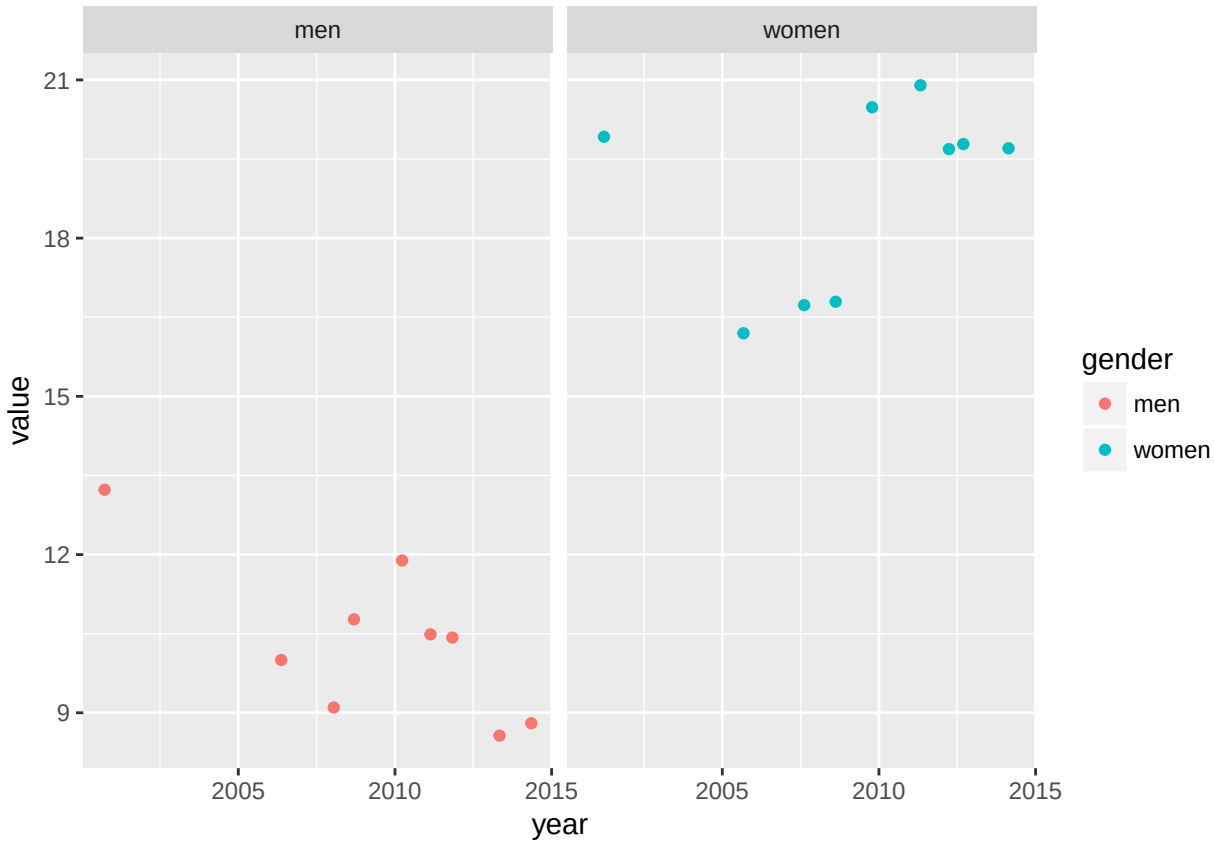
```
## Warning: Ignoring unknown aesthetics: xmax
```



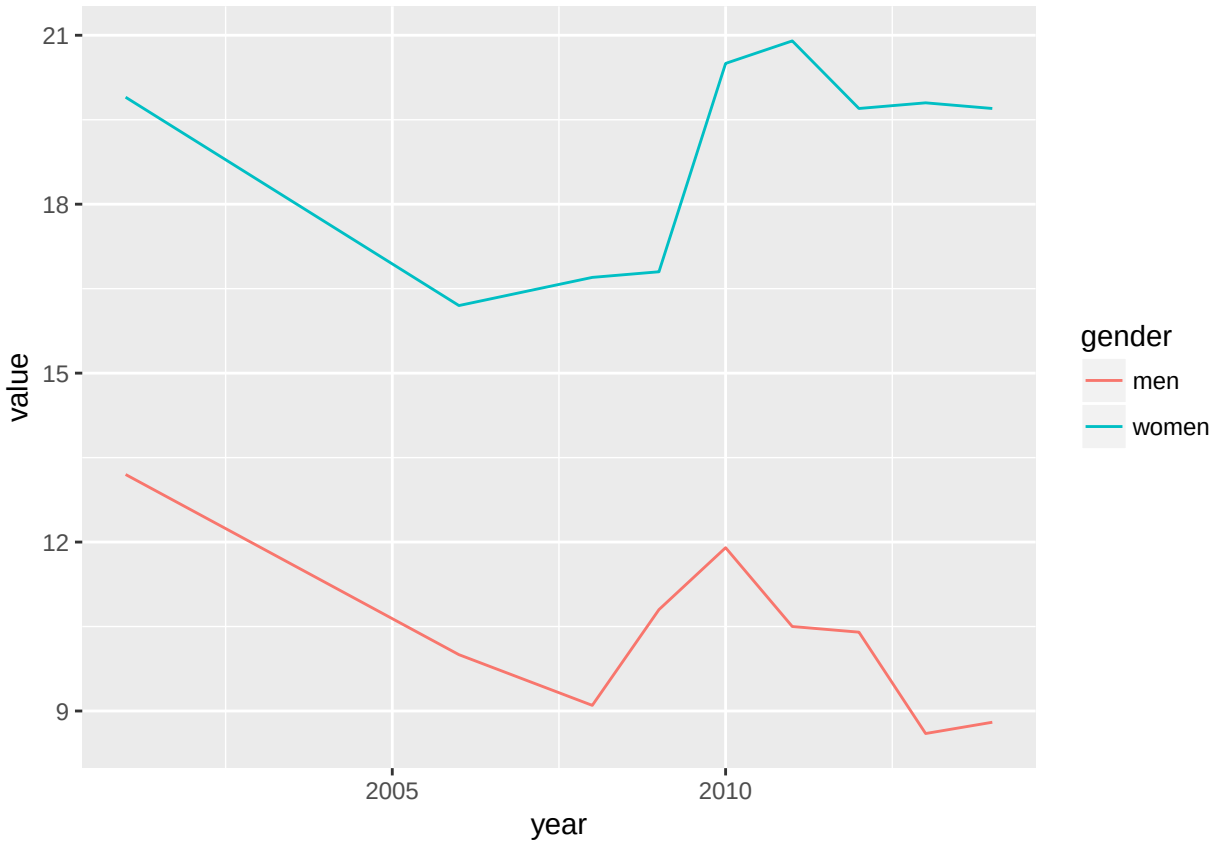

```
#creating geom bar with fill for gender and the psotion of dodge

ggplot(joint_unemployment_total,
       aes(x = year,
           y = value,
           fill = gender)) +
  geom_bar(stat = "identity",
           position = "dodge")
```




```
#creating geom gitter with col for gender and facet_grid for gender
ggplot(joint_unemployment_total, aes(x = year, y = value, col = gender)) +
  geom_jitter() +
  facet_grid(. ~ gender)
```



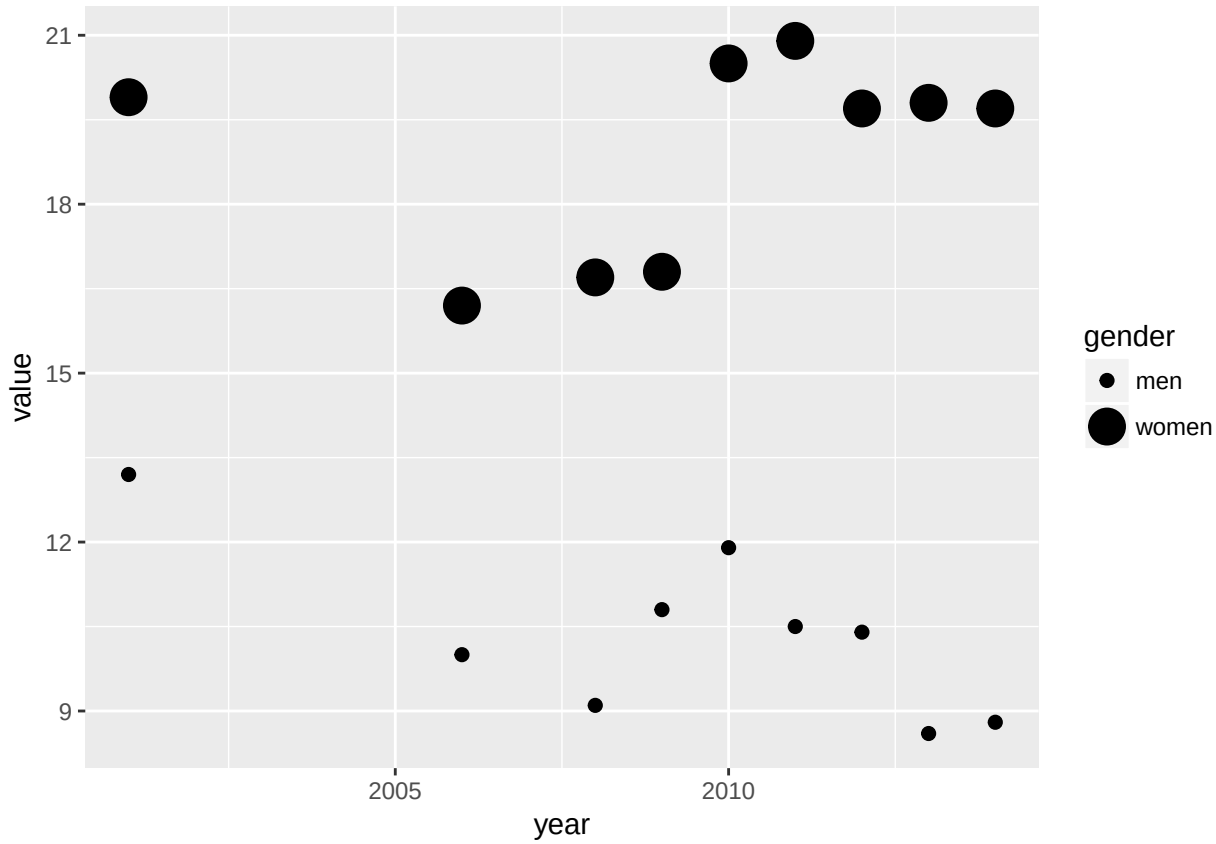

```
#creating geom line with color for gender
ggplot(joint_unemployment_total,
       aes(x = year,
           y = value,
           color = gender)) +
  geom_line()
```



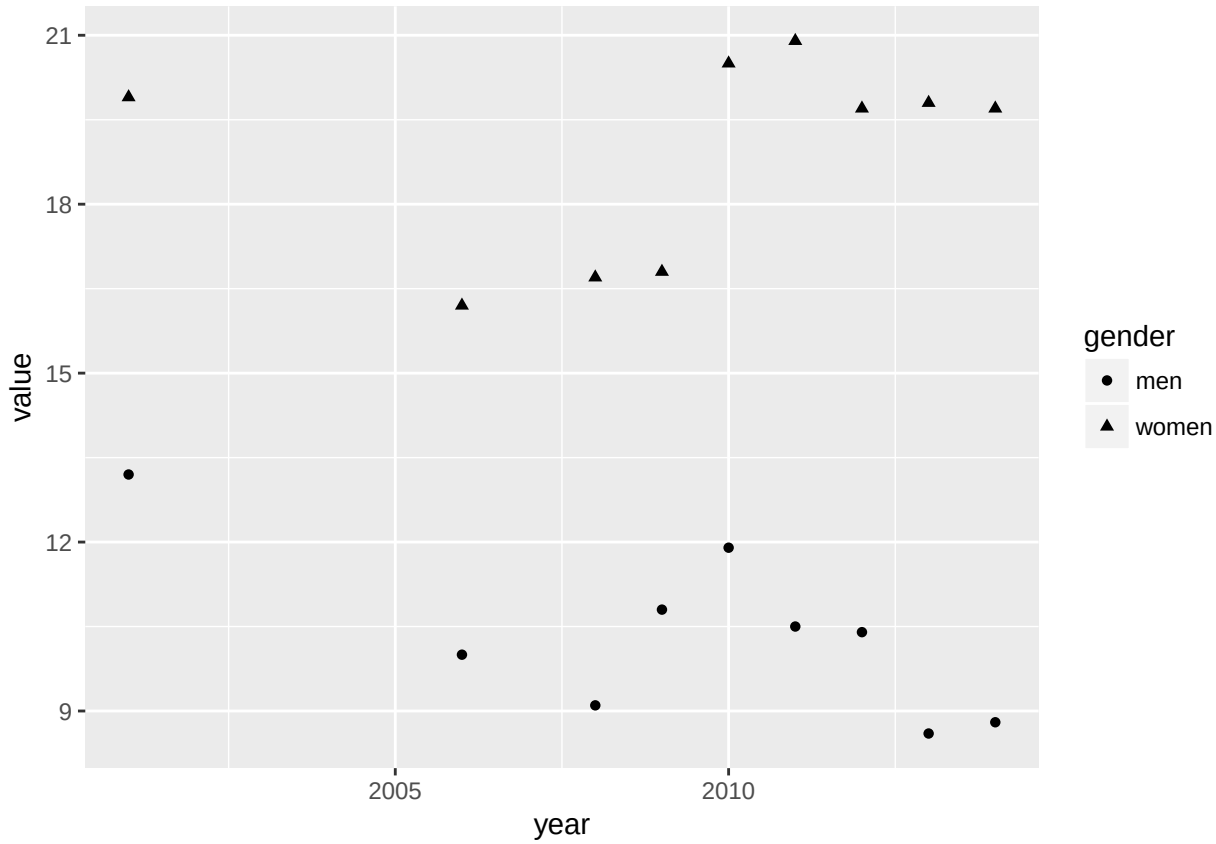
4. Create at least three variations of the plots you've already made by modifying some of the arguments we covered in class (i.e. `position`, `scale`, `size`, `linetype` etc.). Do any of these modifications help you understand your data better? Why or why not? Do any of them create a misleading interpretation of the relationships between your variables? If yes, how so?

```
# maniplating the size for gender
ggplot(joint_unemployment_total,
       aes(x = year,
           y = value,
           size = gender)) +
  geom_point()
```

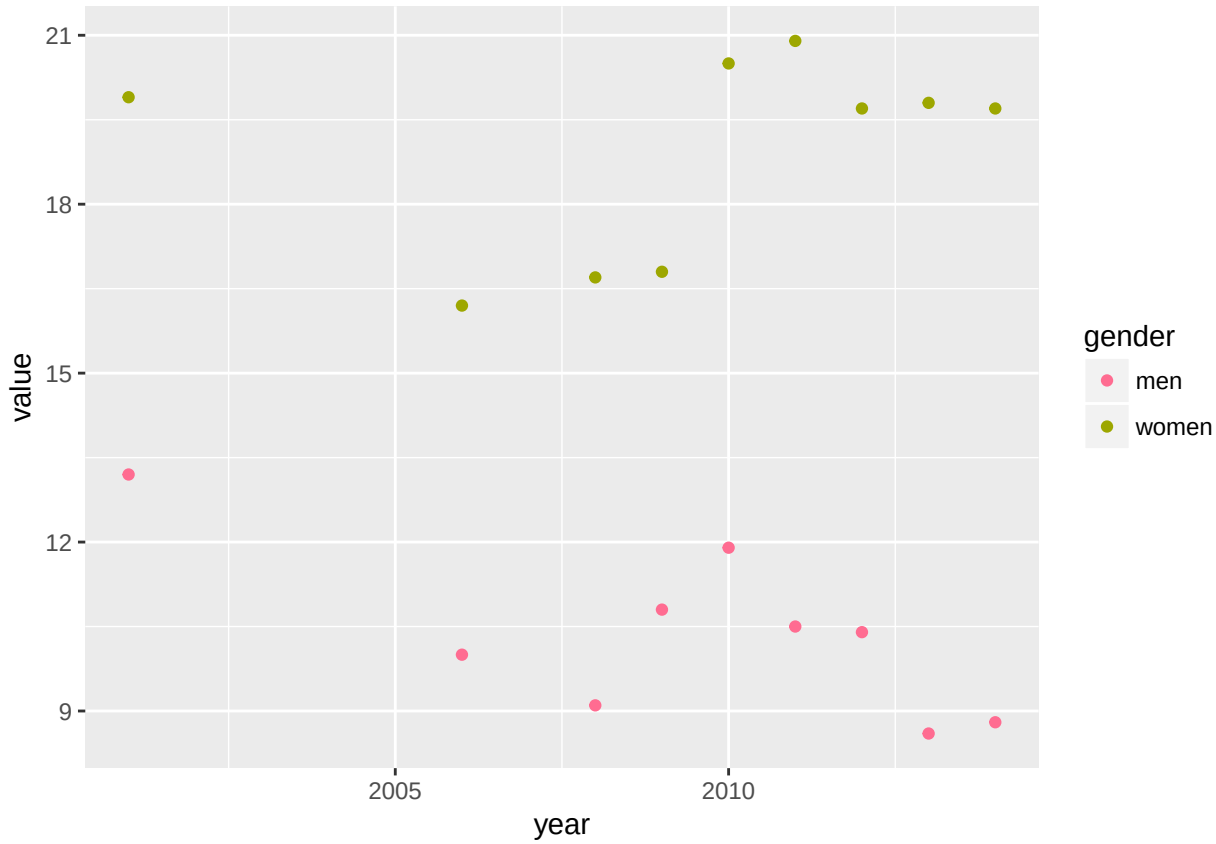
```
## Warning: Using size for a discrete variable is not advised.
```



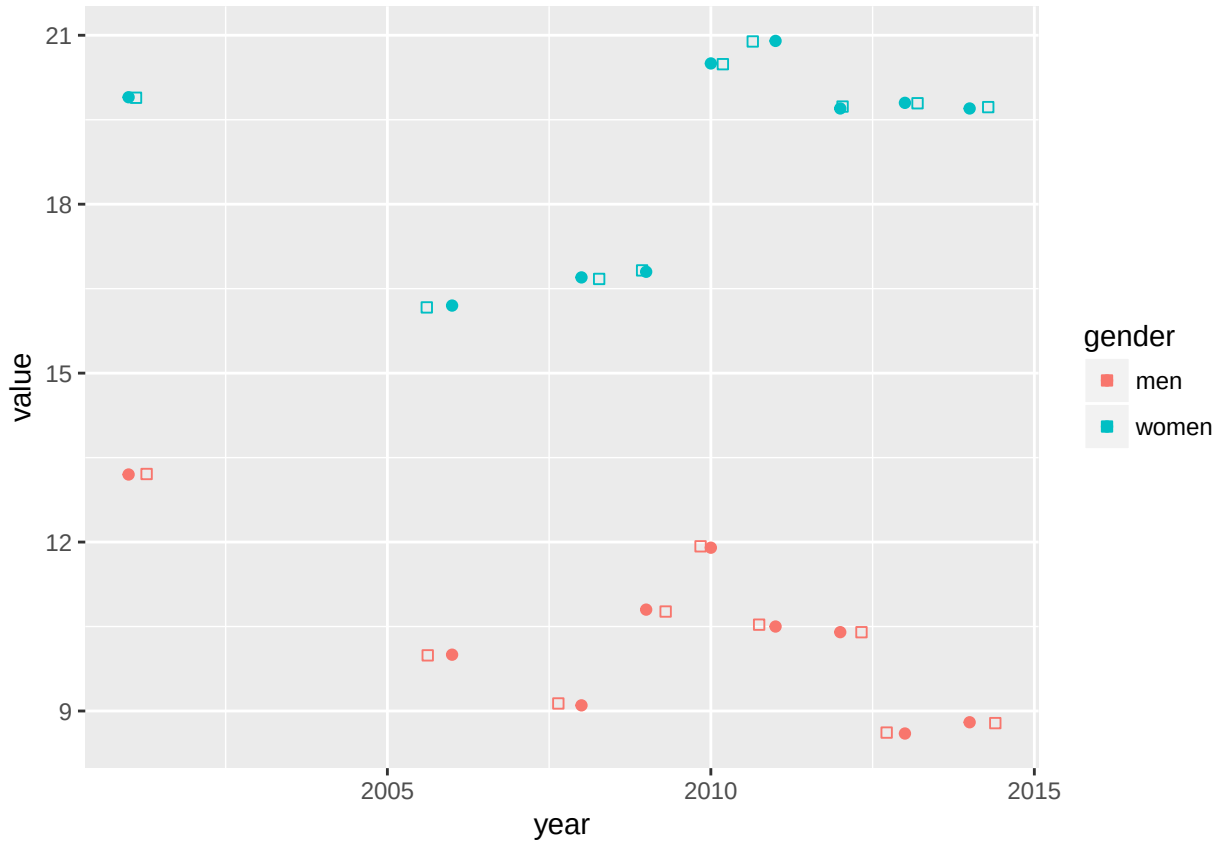

```
# manipulating the shape for gender
ggplot(joint_unemployment_total,
       aes(x = year,
           y = value,
           shape = gender)) +
  geom_point()
```



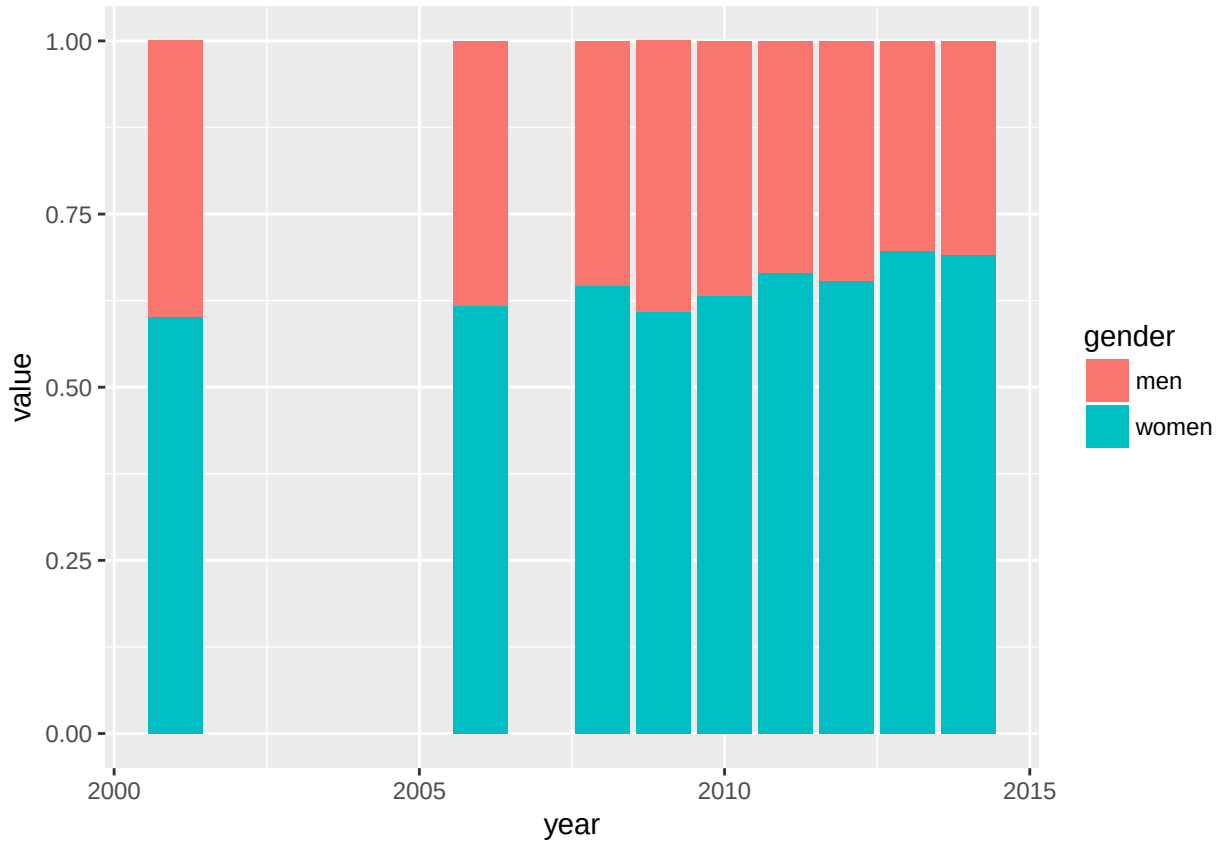

```
# adding scale_colour_hue
ggplot(joint_unemployment_total,
       aes(x = year,
           y = value,
           col = gender)) +
  geom_point() +
  scale_colour_hue(h = c(0, 90))
```



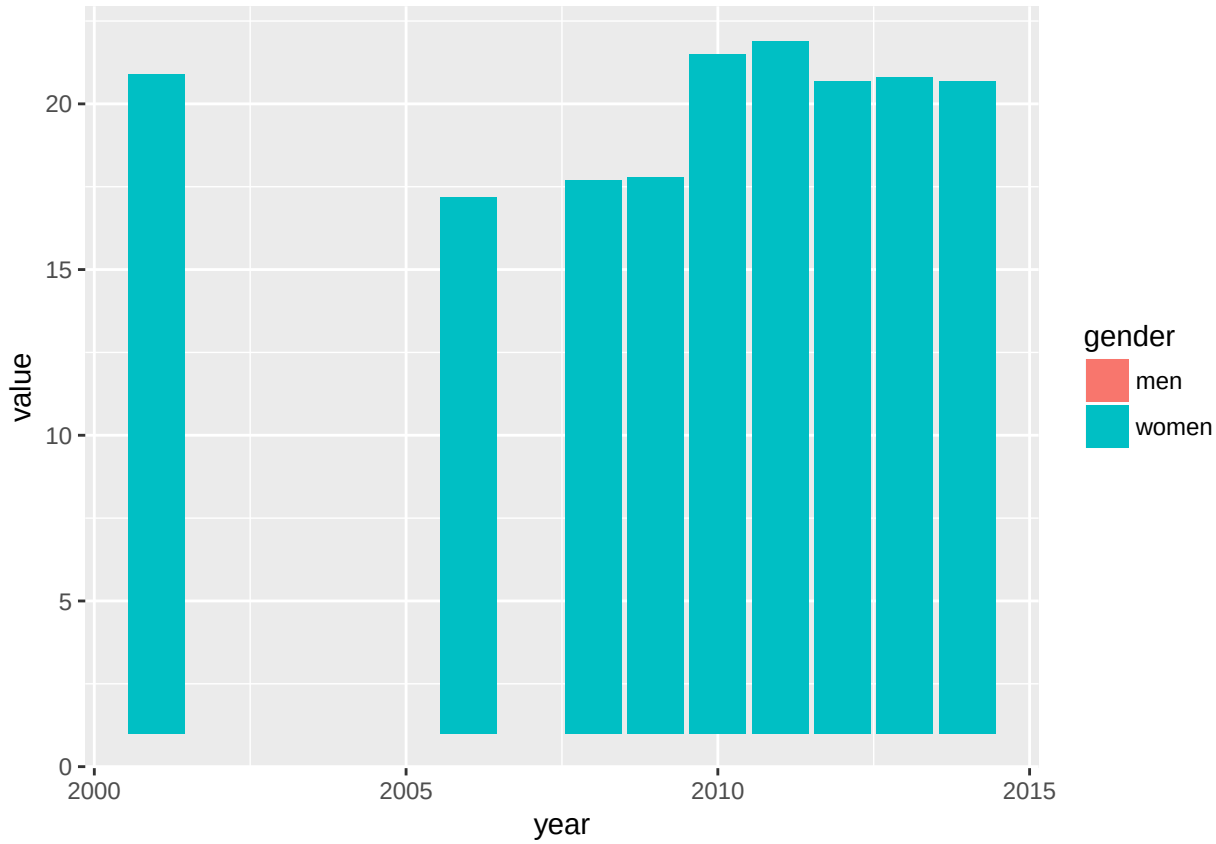

```
#praticing with shape 0 of geom jitter

ggplot(joint_unemployment_total,
       aes(x = year,
           y = value,
           col = gender)) +
  geom_point() +
  geom_jitter(shape = 0)
```



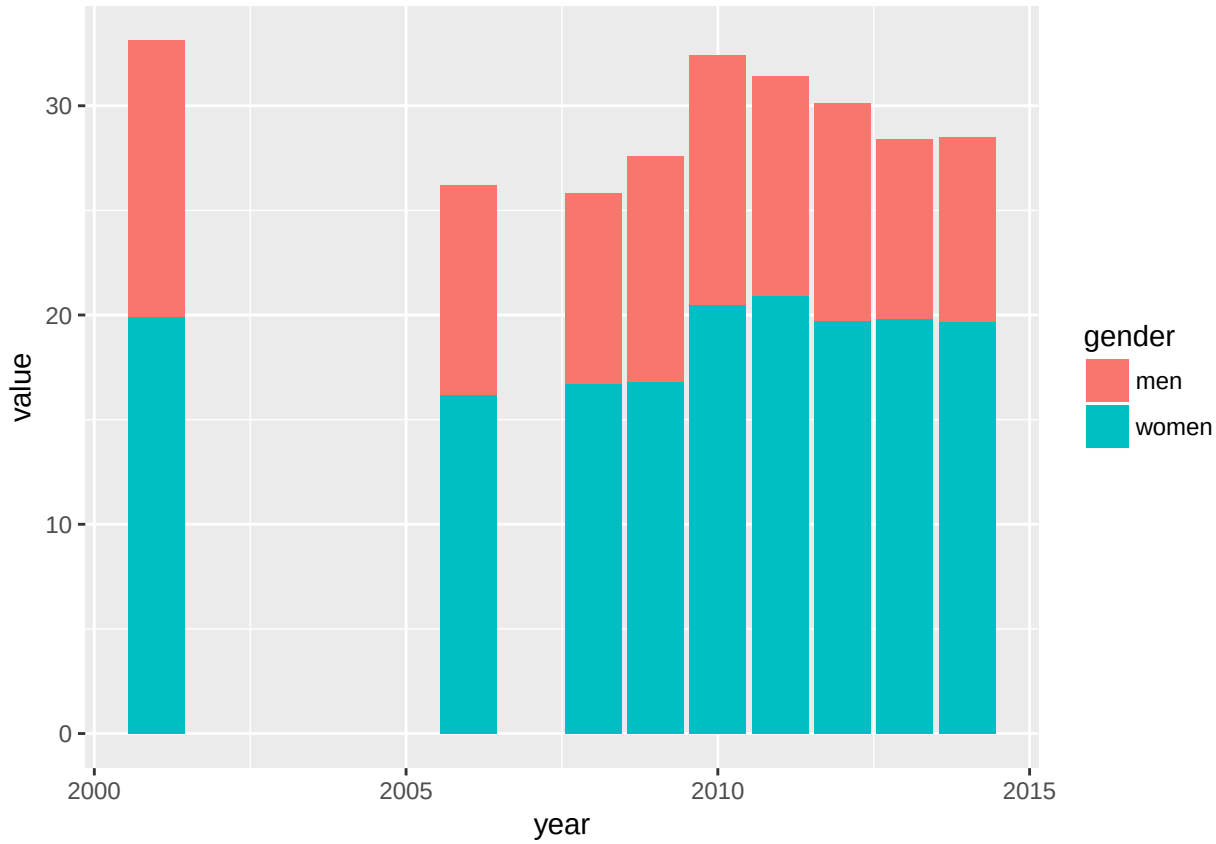

```
# geom bar with the state of identity and position of fill
ggplot(joint_unemployment_total,
       aes(x = year,
           y = value,
           fill = gender)) +
  geom_bar(stat = "identity",
           position = "fill")
```




```
# creating geom bar with  position_nudge

ggplot(joint_unemployment_total,
       aes(x = year,
           y = value,
           fill = gender)) +
  geom_bar(stat = "identity",
           position = position_nudge(y = 1))
```

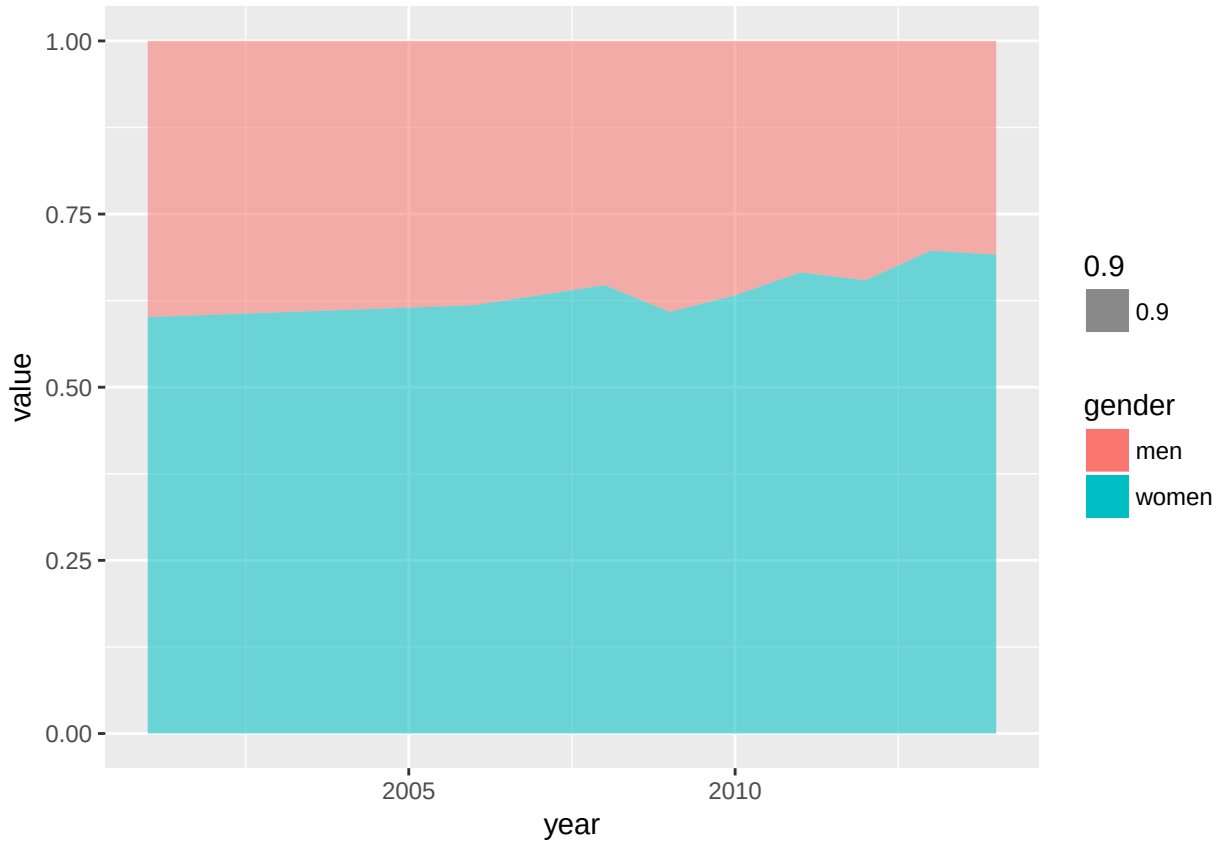



```
# geom bar with the position of stack
ggplot(joint_unemployment_total,
       aes(x = year,
           y = value,
           fill = gender)) +
  geom_bar(stat = "identity",
           position = "stack")
```

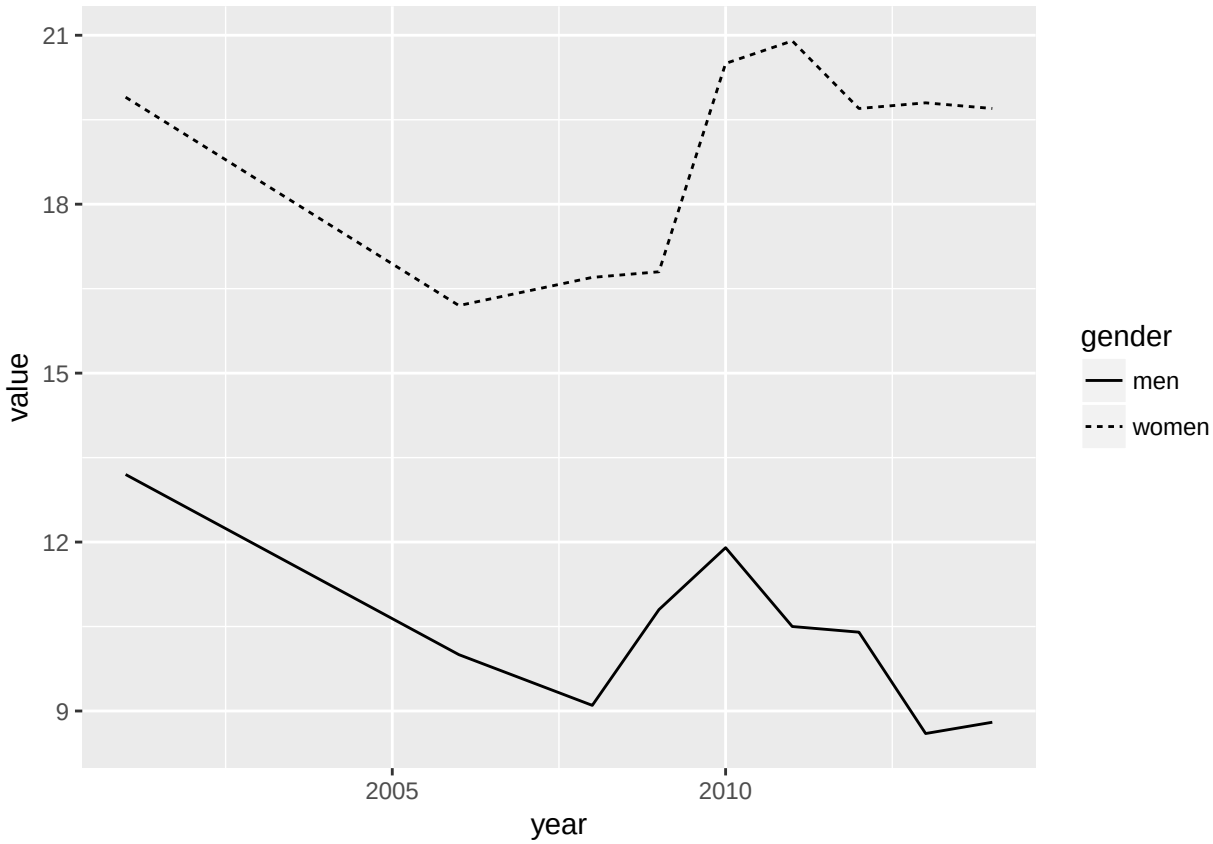



```
# geom ribbon with the position of fill
ggplot(joint_unemployment_total,
       aes(x = year,
           y = value,
          fill = gender)) +
  geom_ribbon(aes(ymax = value, xmax = year,
                  ymin = 0,
                  alpha = 0.9),
              position = "fill")
```

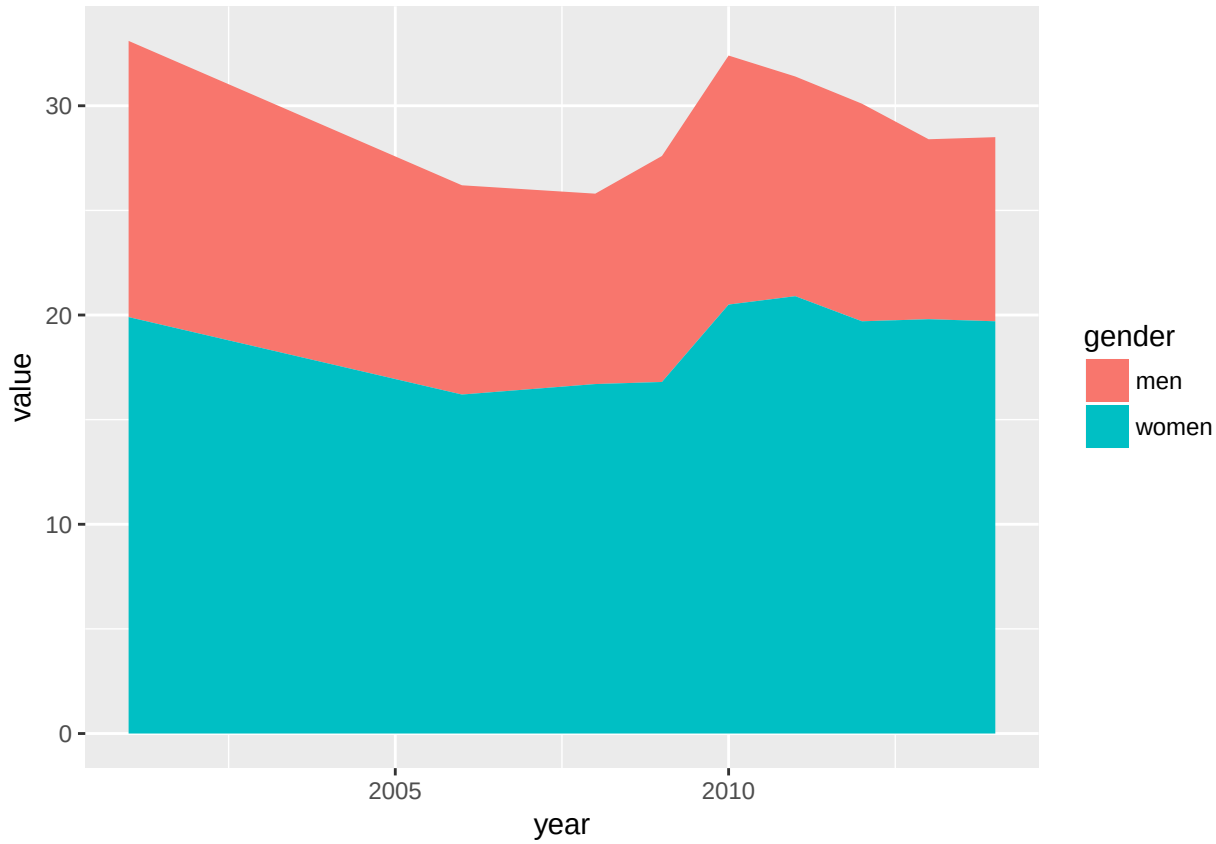
```
## Warning: Ignoring unknown aesthetics: xmax
```



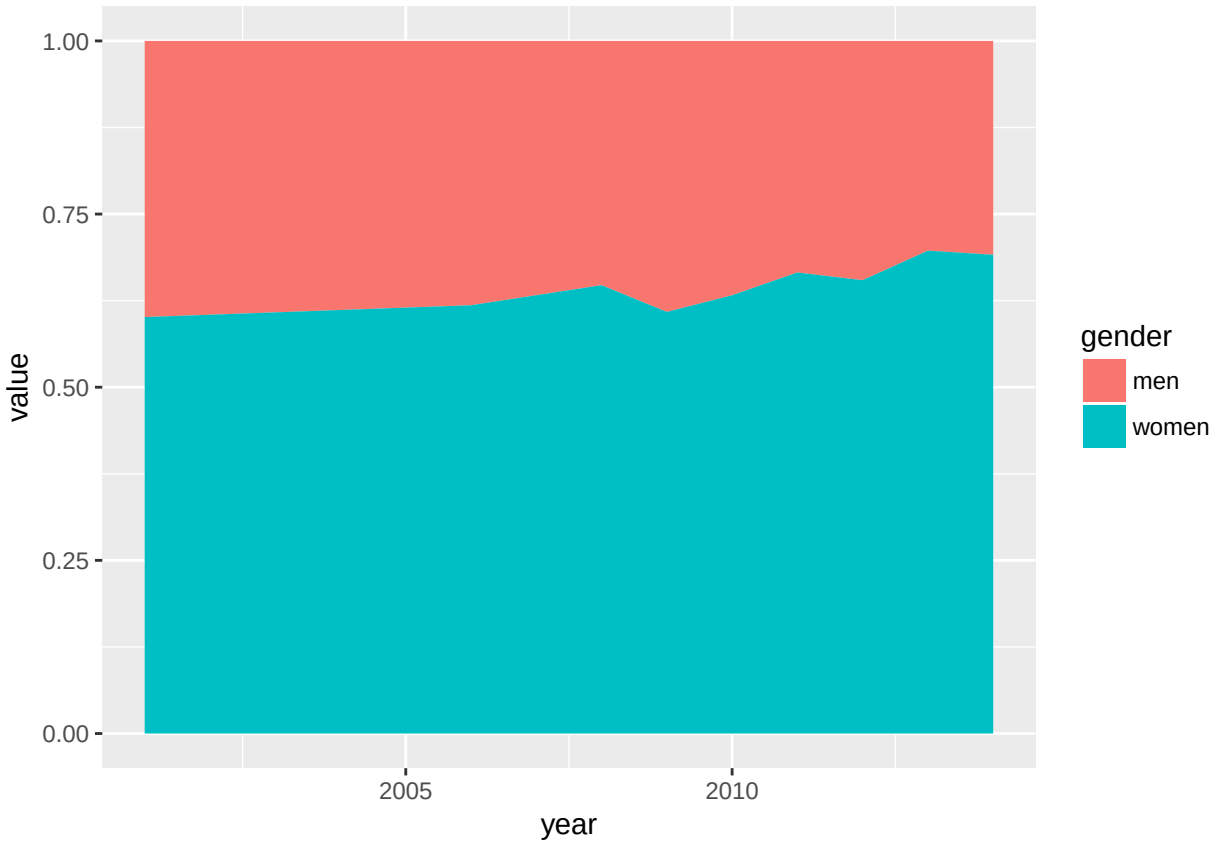

```
# geom line with linetype for gender
ggplot(joint_unemployment_total,
       aes(x = year,
           y = value,
           linetype = gender)) +
  geom_line()
```




```
#geom area with fill for gender
ggplot(joint_unemployment_total,
       aes(x = year,
           y = value,
           fill = gender)) +

  geom_area()
```



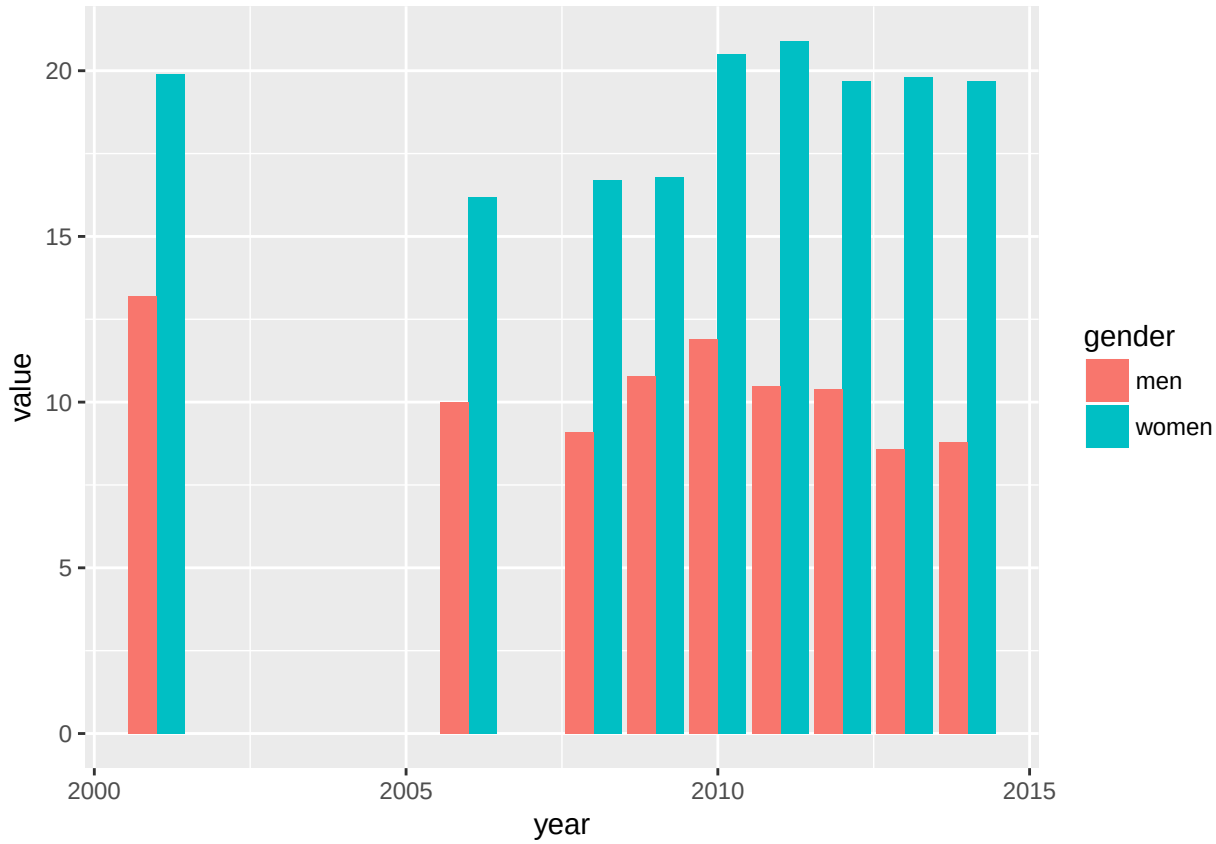

```
# geom_area(position = "fill") and fill for gender
ggplot(joint_unemployment_total,
       aes(x = year,
           y = value,
           fill = gender)) +

  geom_area(position = "fill")
```



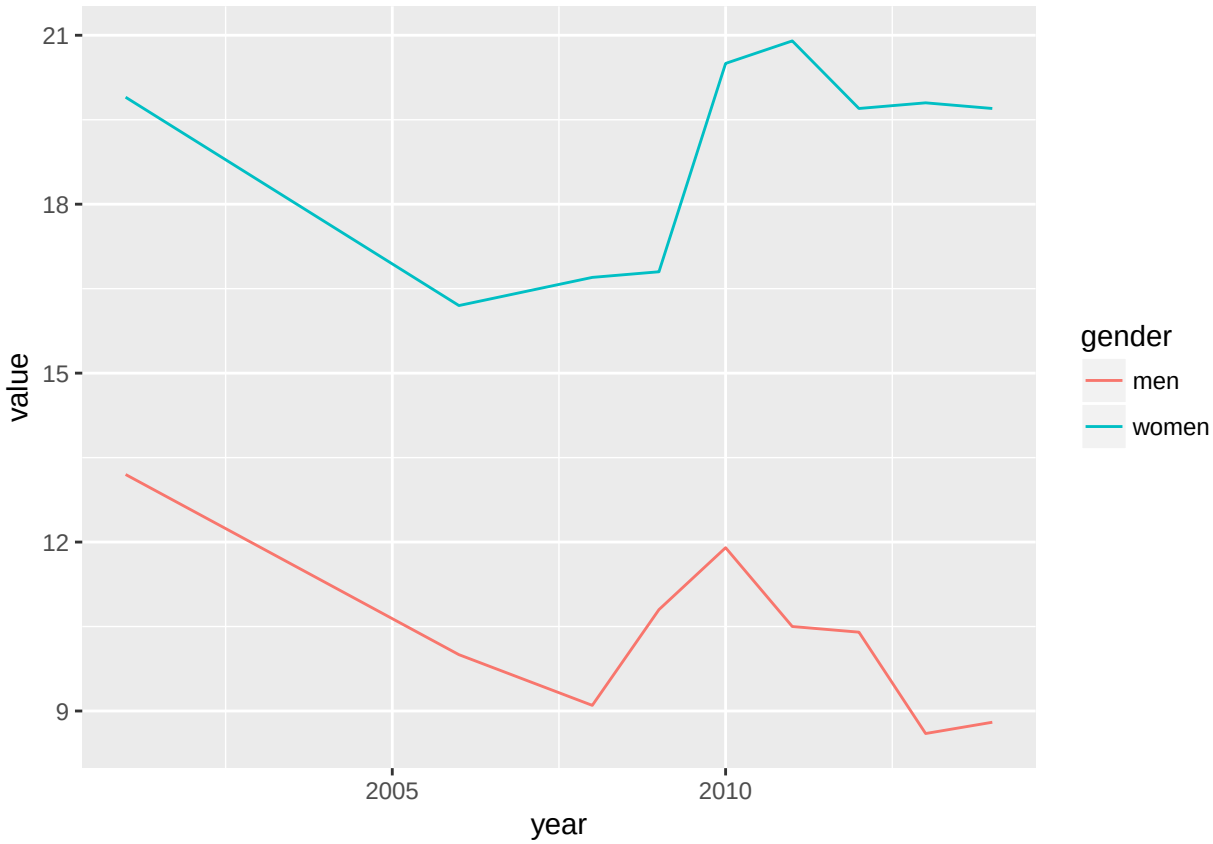
5. From the plots you've created thus far, do any of them seem appropriate for a general audience? Why or why not? If so, what do you think you'd still need to do to make them more suitable as explanatory visualizations?

Yes, I think the two plots which I created below are more appropriate for general audience. Both show the unemployment rate by years for men and women. Comparing is easy in these plot. They are in simple form, so the audience can focus on the data not the merely visualization of data. For both these plots a title is required to indicate that the plots reflect unemployment rate for two genders in Iran.

```
# the fist appropriate plot for general audiance
ggplot(joint_unemployment_total,
       aes(x = year,
           y = value,
           fill = gender)) +
  geom_bar(stat = "identity",
           position = "dodge")
```




```
# the second appropriate plot for general audiance
ggplot(joint_unemployment_total,
       aes(x = year,
           y = value,
           color = gender)) +
  geom_line()
```


21
18
15
12
9
value
2005
2010
year
gender
men
women