

ggplot 2 Lab

Daisy Cutter

2025-09-19

```
#read the csv and call the data set my_nile
my_nile <- read.csv('https://vinnys-classes.github.io/data/Nile.csv')

#look at the first six rows using head() and the last six using tail() to make sure it worked like we expected and to see what the data looks like.
head(my_nile)
```

	flow <int>	year <int>	changepoint <int>
1	1120	1871	0
2	1160	1872	0
3	963	1873	0
4	1210	1874	0
5	1160	1875	0
6	1160	1876	0
6 rows			

```
tail(my_nile)
```

	flow <int>	year <int>	changepoint <int>
95	912	1965	1
96	746	1966	1
97	919	1967	1
98	718	1968	1
99	714	1969	1
100	740	1970	1
6 rows			

#finally, we need to make sure the package is loaded. Note there isn't any quotations when we reference the package with library()

```
library(ggplot2)
```

#HINT: if you get an error that reads there is no package called 'ggplot2' then please run, in your console, the code install.packages('ggplot2') You only have to run that line once so don't put it into your markdown file.

#You don't need to know the following comment/understand the code but I wanted to give you a more full picture.

#Finally we need to change the class of a variable. Currently R has changepoint as an integer and we need it to be either logical (TRUE/FALSE) or categorical which R calls a factor. The use of as.????() is a common way to change classes in R (eg as.numeric() or as.logical())

```
class(my_nile$changepoint)
```

```
## [1] "integer"
```

```
my_nile$changepoint <- as.factor(my_nile$changepoint)
class(my_nile$changepoint)
```

```
## [1] "factor"
```

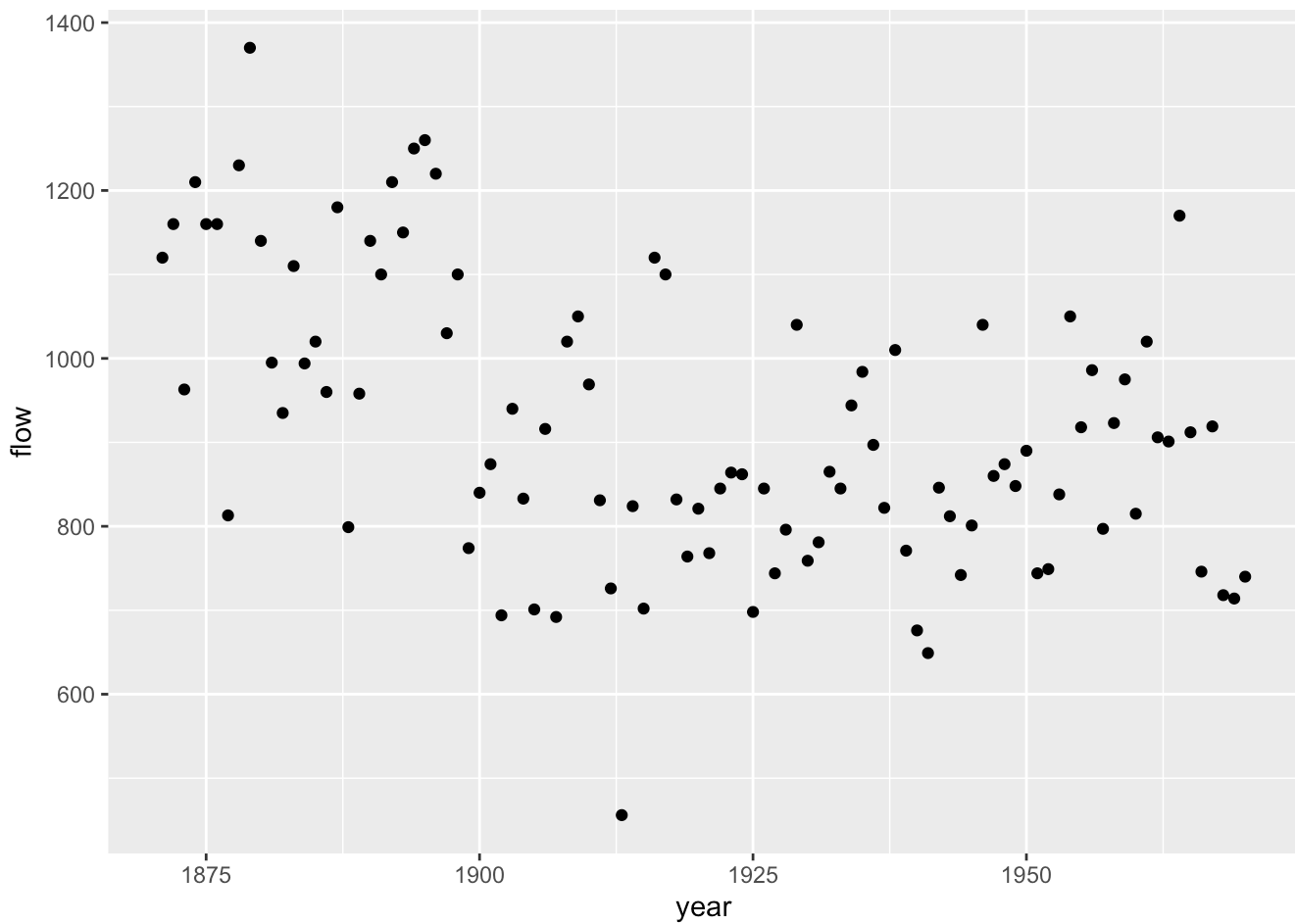
Question 1

Please make a new Rmd file and indicate the names and types of variables we have in our data.

We have 3 variables: flow, year, and changepoint. Flow and year are integers, changepoint is categorical with a 1 or a 0.

Question 2

```
ggplot(data = my_nile,
       mapping = aes(x = year,
                     y = flow)) +
  geom_point()
```



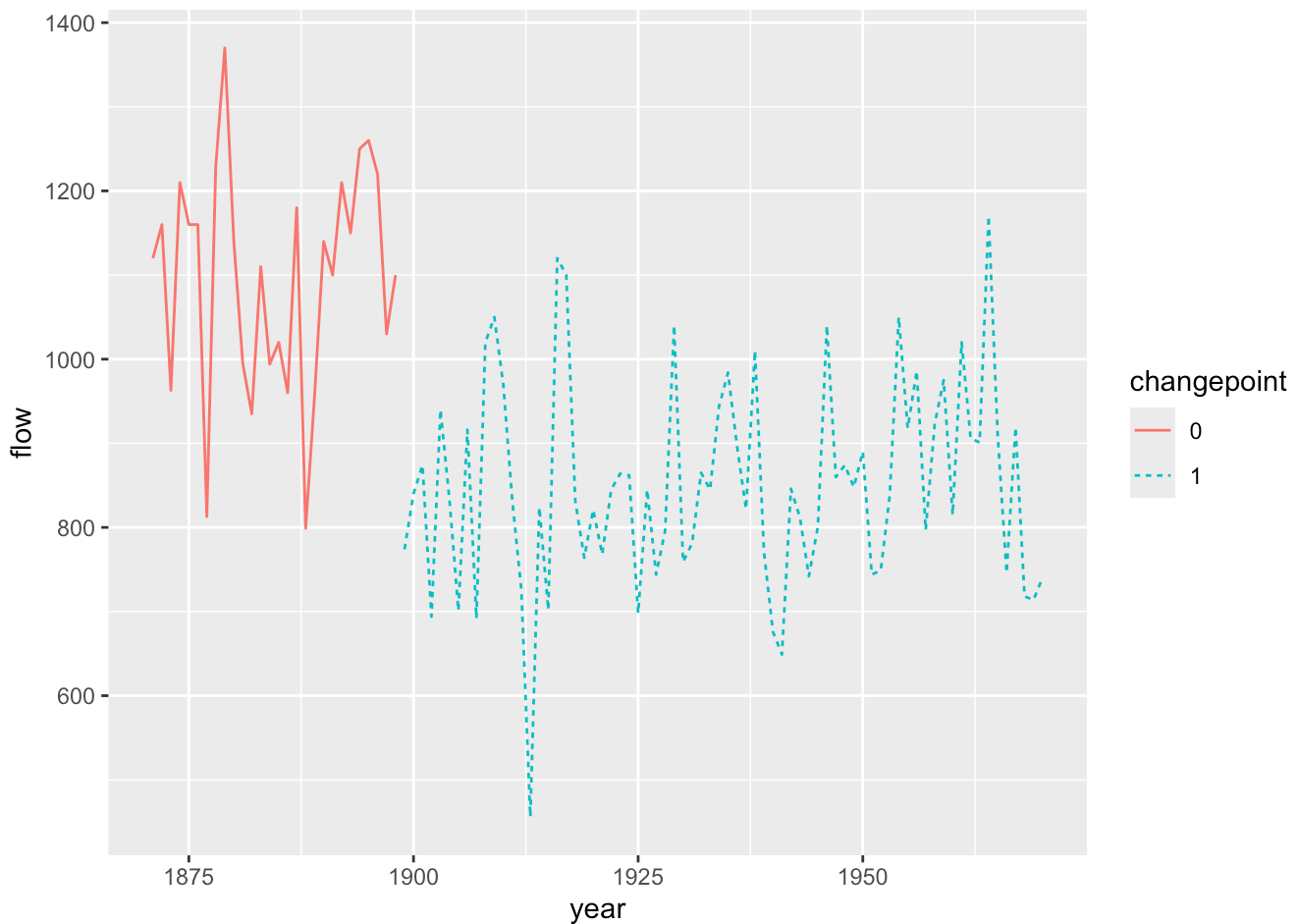
Question 3

I prefer the original graph that we are trying to recreate over the scatterplot, it is easier to tell how the flow changes by year than with a scatterplot. The graph above is hard to read.

Question 4

```
ggplot(data = my_nile,
       mapping = aes(x = year,
                     y = flow)) +

  geom_line(mapping = aes(color = changepoint,
                          linetype = changepoint))
```



Question 4

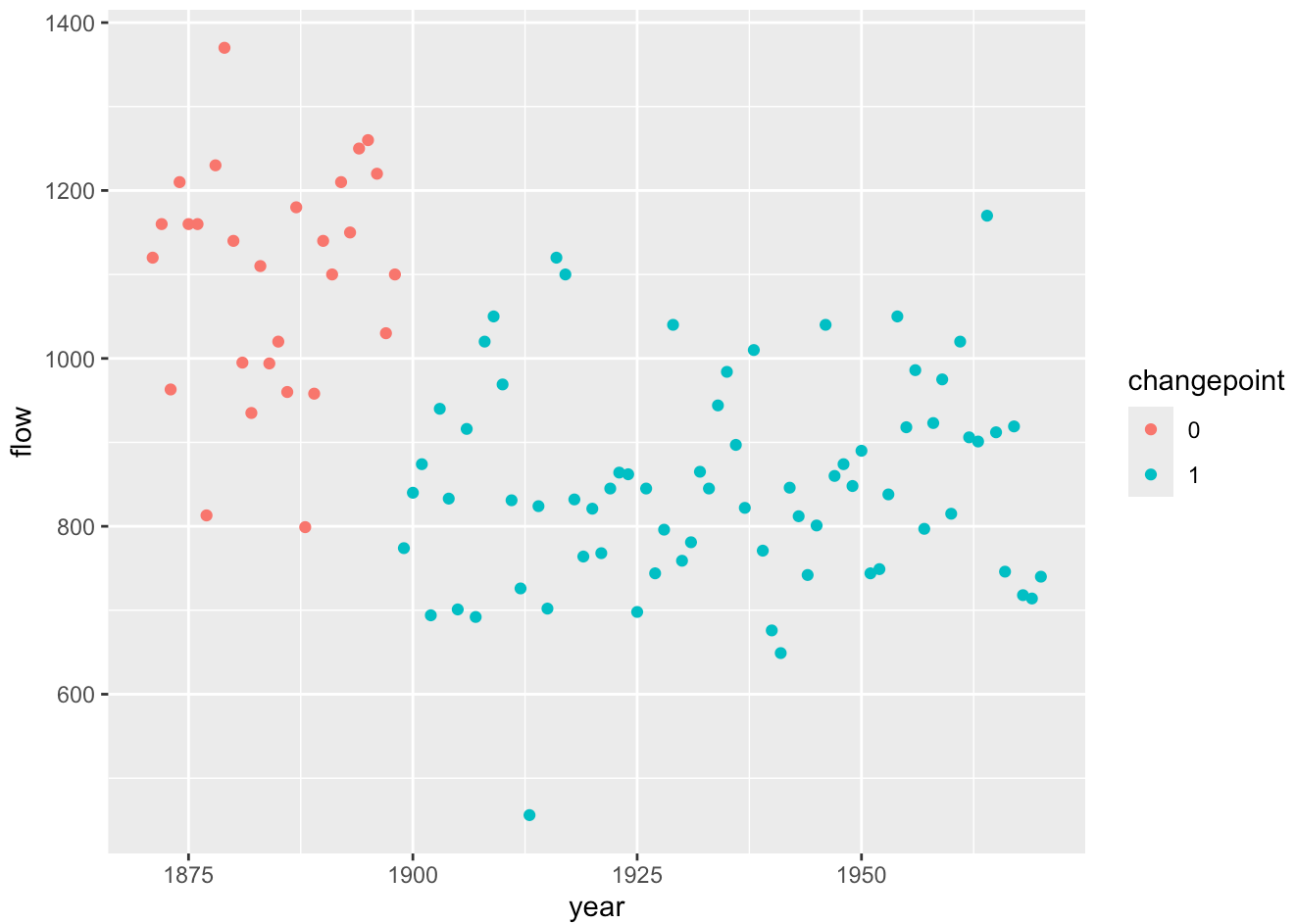
Please discuss the difficulty needed to add line type in addition to color in order to make the graph color blind friendly.

There needs to be a differentiation on the graph rather than just color, which is why the linetype adding a texture to the lines is useful. It makes it easier to distinguish what is happening, even if you cannot distinguish the colors, you can tell that one line is smooth and one is dotted. It is not much harder, you just need to know what linetype is and how to use it.

Question 5

Let's go back to the scatterplot idea from Question 2. Please remake that graph (feel free to copy and paste) but this time add **just** color to the graph.

```
ggplot(data = my_nile,
       mapping = aes(x = year,
                     y = flow)) +
  geom_point(mapping = aes(color=changepoint))
```



Question 6

Because it's points and not lines the parameter `linetype` won't be useful. Instead, the parameter we'd like to set in its place is called `shape`. Please do this now.

```
ggplot(data = my_nile,  
       mapping = aes(x = year,  
                     y = flow)) +  
  geom_point(mapping = aes(color = changepoint,  
                          shape = changepoint))
```



Question 7

Using any of the graphs made so far (with strong preference to the scatterplot), please describe the relationship between years and flow rate. You do not need to discuss the change point; I'm wanting you to practice describing a scatterplot. Mention the form, direction, strength, and outliers please.

On this scatterplot, there is a medium linear association between year and flow rate, as the year increases the flow rate tends to go down. There is an outlier where the flow is very low at around 1911, and very high around 1950. There is also an outlier where the flow rate is very high shortly after 1875.

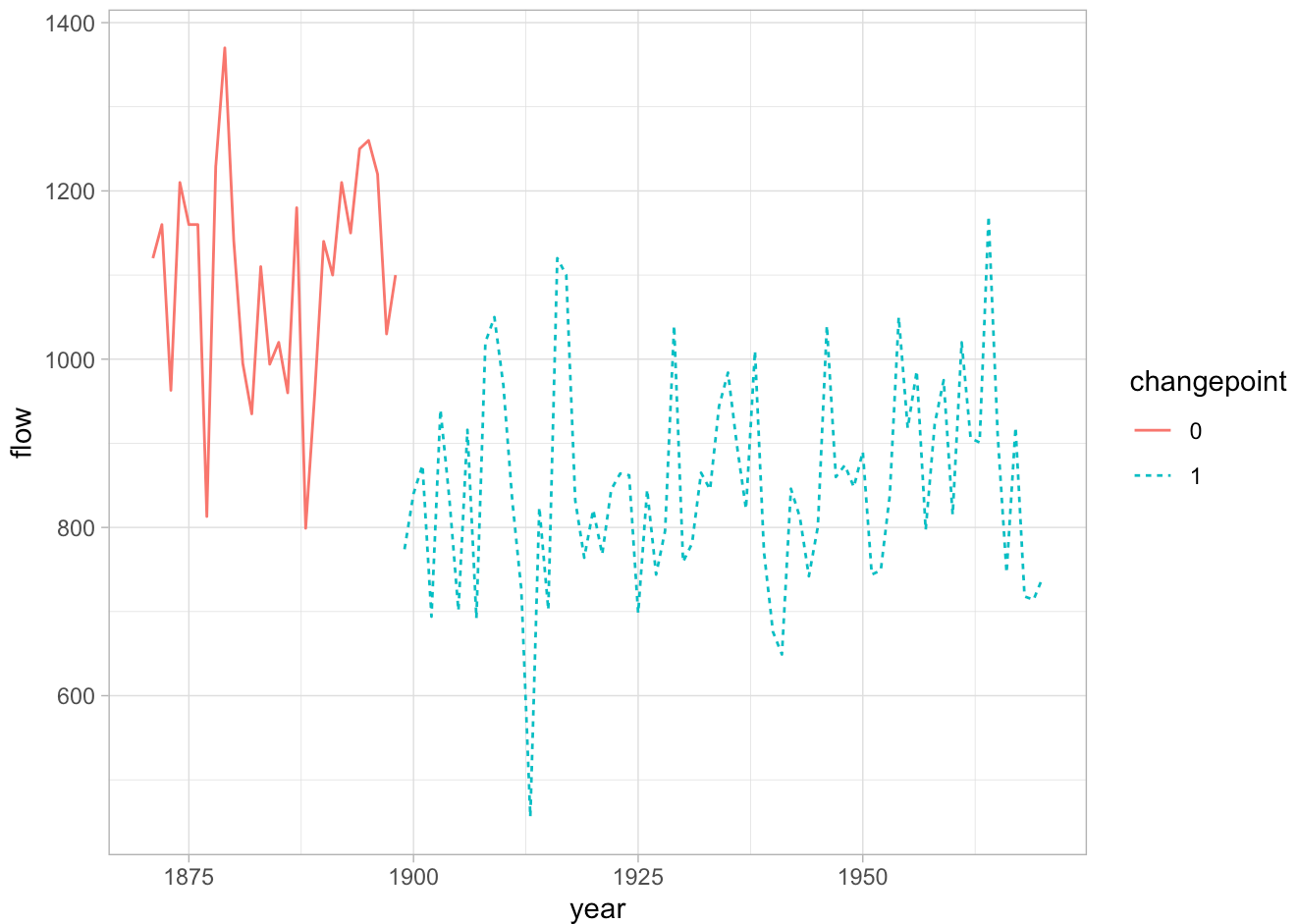
Question 8

I like theme_light, as I think that balances the most readability with appearance. It is more readable than theme_minimal as it has grid lines, and is more readable than theme_linedraw because it has lighter grid lines.

```
ggplot(data = my_nile,
       mapping = aes(x = year,
                     y = flow)) +

  geom_line(mapping = aes(color = changepoint,
                          linetype = changepoint)) +

  theme_light()
```



Question 9 & Question 10

```
nile_plot <- ggplot(data = my_nile,
  mapping = aes(x = year,
    y = flow)) +

  geom_line(mapping = aes(color = changepoint,
    linetype = changepoint)) +

  theme_light(base_size = 15) +
  xlab("Year") +
  ylab("Flow rate") +

  labs(title = 'The Nile Rivers Flow Over Time',
    subtitle = 'Daisy Cutter',
    color = 'Changepoint',
    linetype = 'Changepoint',
    alt = 'A graph showing the flow of the Nile river from 1871 to 1970, showing that
    before about 1900, the flow of the Nile River was much higher. There is a sharp drop-of
    f, and flow drops beyond this period consistently.')

get_alt_text(nile_plot)
```

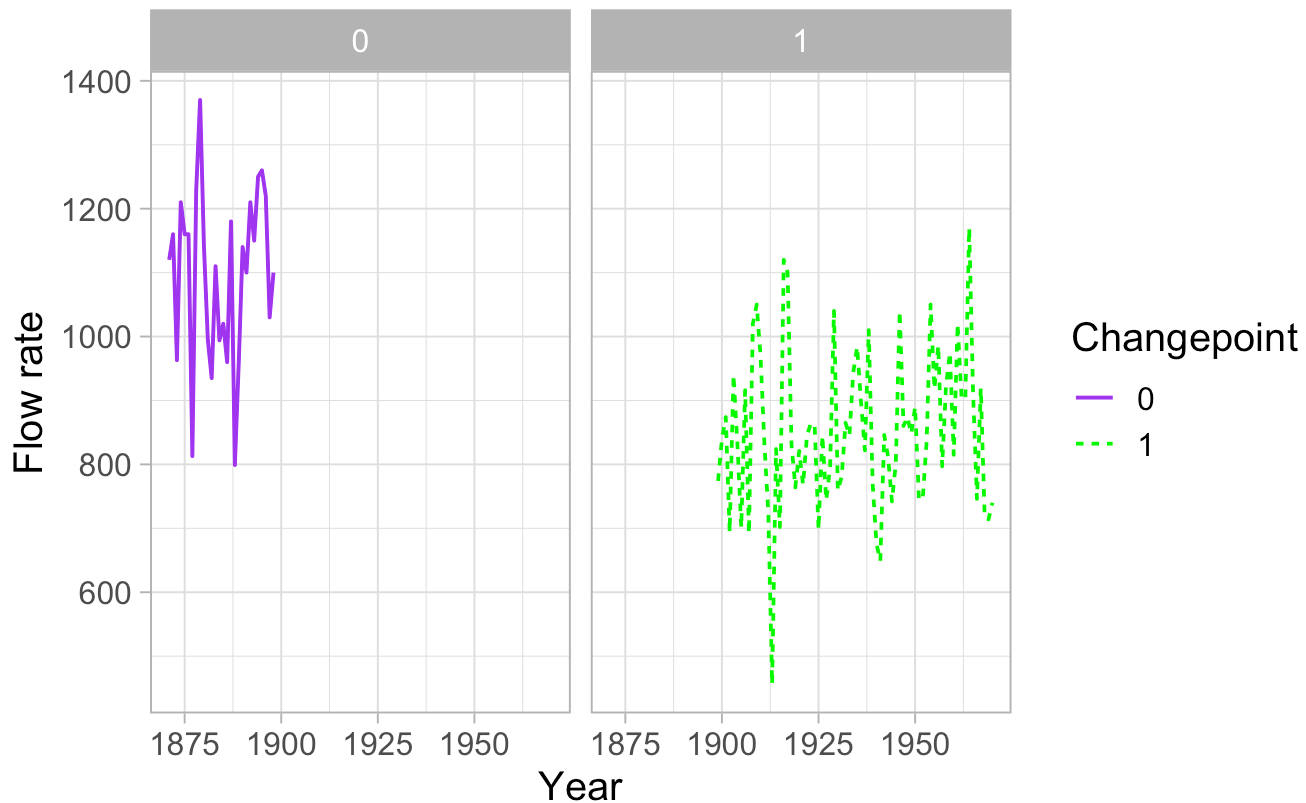
```
## [1] "A graph showing the flow of the Nile river from 1871 to 1970, showing that before about 1900, the flow of the Nile River was much higher. There is a sharp drop-off, and flow drops beyond this period consistently."
```

Facets Section

```
ggplot(data = my_nile,  
       mapping = aes(x = year,  
                     y = flow)) +  
  
  geom_line(mapping = aes(color = changepoint,  
                         linetype = changepoint)) +  
  
  theme_light(base_size = 15) +  
  xlab("Year") +  
  ylab("Flow rate") +  
  
  labs(title = 'The Nile Rivers Flow Over Time',  
       subtitle = 'Daisy Cutter',  
       color = 'Changepoint',  
       linetype = 'Changepoint',  
       alt = 'A graph showing the flow of the Nile river from 1871 to 1970, showing that before about 1900, the flow of the Nile River was much higher. There is a sharp drop-off, and flow drops beyond this period consistently.') +  
  facet_grid(~ changepoint) +  
  scale_color_manual(values = c('purple', 'green'))
```

The Nile Rivers Flow Over Time

Daisy Cutter



Question 12

Please explain why there is “blank” space to the right of the purple solid line on the left graph and empty space to the left of the green dashed line.

Because that is the year where the changepoint occurs, on a graph where they are not separated in a facet grid (like the previous ones) it would show a continuous graph.

Question 13

Let's get rid of that blank space and instead make the axis “free” so they can have the most natural axis for the data in each pane. To do this, go to the help page for `facet_grid()` and look at the parameter called “scales”. Choose the option that allows *only* the x-axis to be “free”.

```
ggplot(data = my_nile,
       mapping = aes(x = year,
                     y = flow)) +

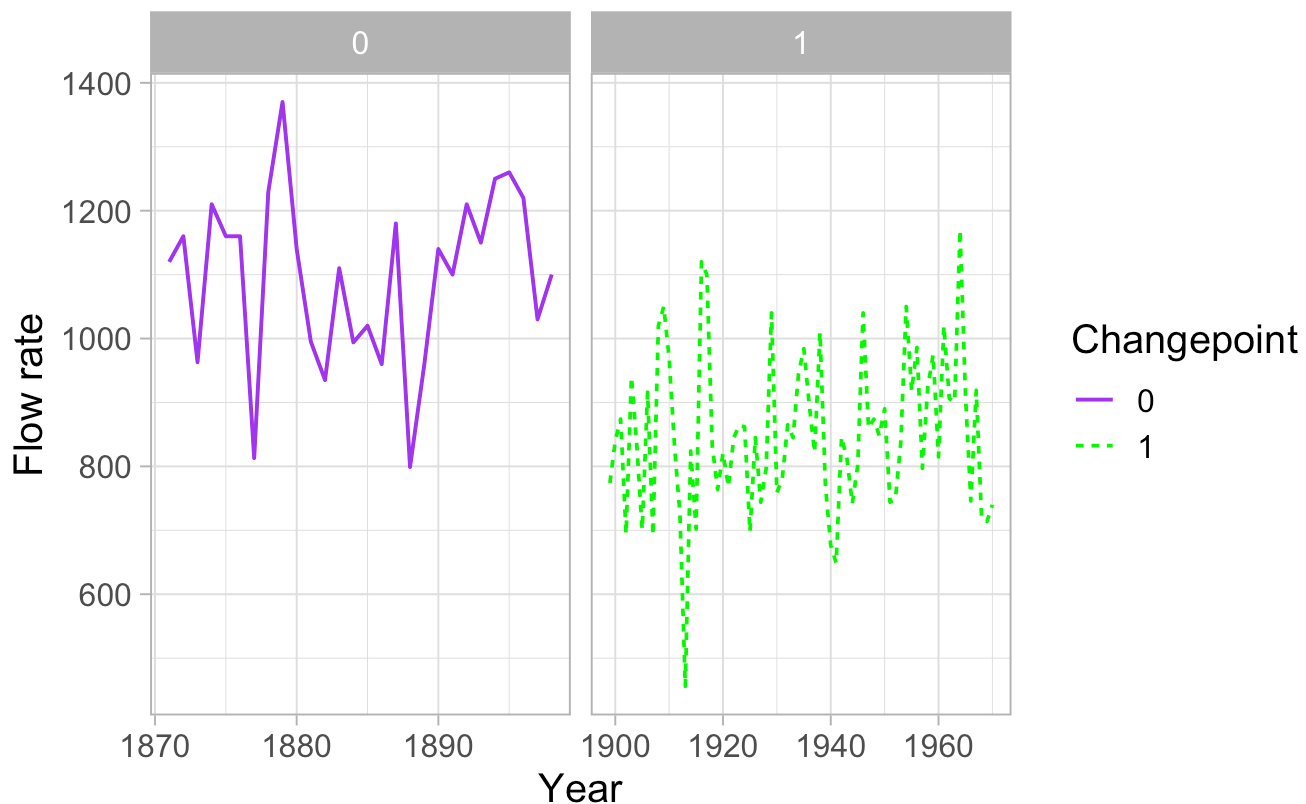
  geom_line(mapping = aes(color = changepoint,
                         linetype = changepoint)) +

  theme_light(base_size = 15) +
  xlab("Year") +
  ylab("Flow rate") +

  labs(title = 'The Nile Rivers Flow Over Time',
       subtitle = 'Daisy Cutter',
       color = 'Changepoint',
       linetype = 'Changepoint',
       alt = 'A graph showing the flow of the Nile river from 1871 to 1970, showing that
before about 1900, the flow of the Nile River was much higher. There is a sharp drop-of
f, and flow drops beyond this period consistently.') +
  facet_grid(~ changepoint, scales = "free_x") +
  scale_color_manual(values = c('purple', 'green'))
```

The Nile Rivers Flow Over Time

Daisy Cutter



Question 14

Copy the graph made above in question 13 and change the side the tilde (the squiggly line) is on from the left side of the “changepoint” variable to the right side. Rerun the graph and comment on what changed.

```

ggplot(data = my_nile,
       mapping = aes(x = year,
                     y = flow)) +

  geom_line(mapping = aes(color = changepoint,
                          linetype = changepoint)) +

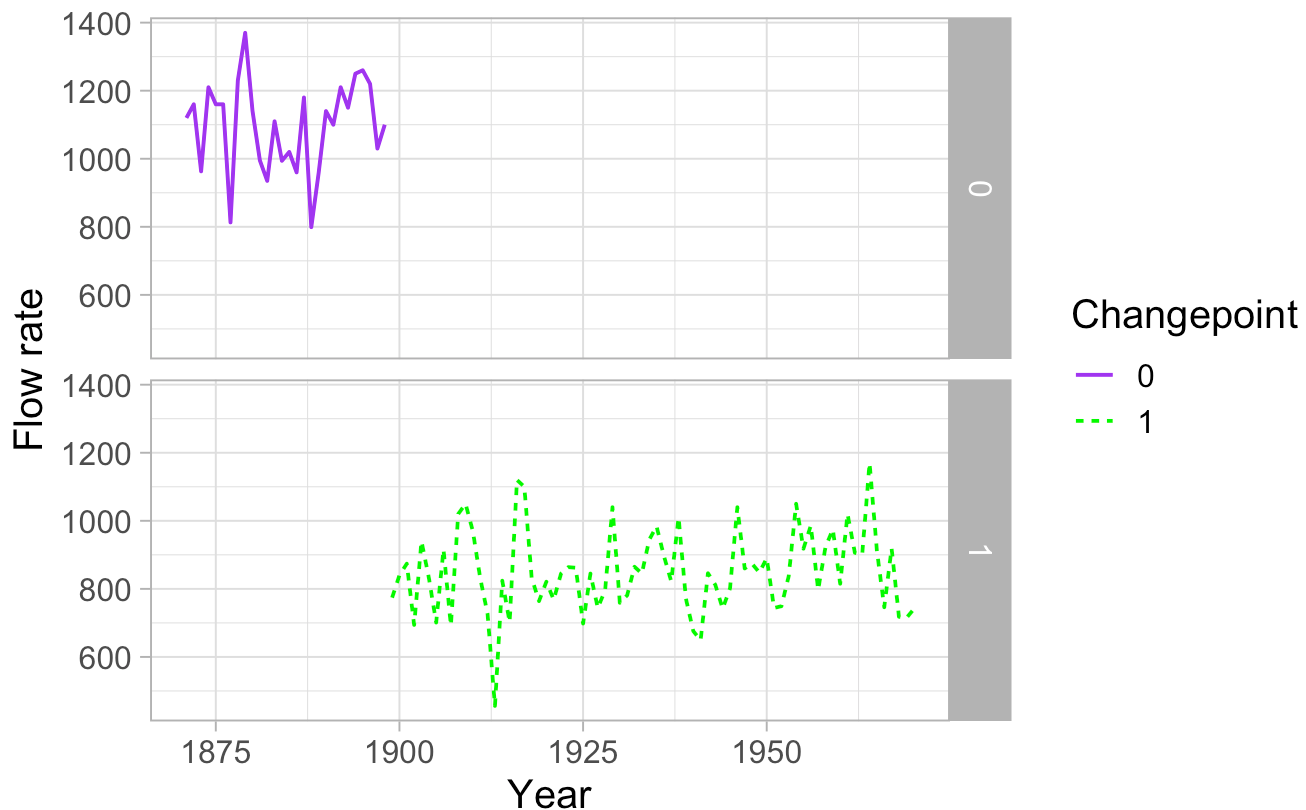
  theme_light(base_size = 15) +
  xlab("Year") +
  ylab("Flow rate") +

  labs(title = 'The Nile Rivers Flow Over Time',
       subtitle = 'Daisy Cutter',
       color = 'Changepoint',
       linetype = 'Changepoint',
       alt = 'A graph showing the flow of the Nile river from 1871 to 1970, showing that
before about 1900, the flow of the Nile River was much higher. There is a sharp drop-of
f, and flow drops beyond this period consistently.') +
  facet_grid(changepoint~., scales = "free_x") +
  scale_color_manual(values = c('purple', 'green'))

```

The Nile Rivers Flow Over Time

Daisy Cutter



The graph now has the two facets on top of each other instead of side to side. The X and Y axes are the same otherwise, but the graphs are stacked vertically instead of horizontally.

Question 15

Above the code that changes the color of the graph to purple and green is the `scale_color_manual()`. Feel free to change the colors to whatever you prefer. You can use `colors()` to see all your color choices.

```
ggplot(data = my_nile,
       mapping = aes(x = year,
                     y = flow)) +

  geom_line(mapping = aes(color = changepoint,
                         linetype = changepoint)) +

  theme_light(base_size = 15) +
  xlab("Year") +
  ylab("Flow rate") +

  labs(title = 'The Nile Rivers Flow Over Time',
       subtitle = 'Daisy Cutter',
       color = 'Changepoint',
       linetype = 'Changepoint',
       alt = 'A graph showing the flow of the Nile river from 1871 to 1970, showing that
before about 1900, the flow of the Nile River was much higher. There is a sharp drop-of
f, and flow drops beyond this period consistently.') +
  facet_grid(~changepoint, scales = "free_x") +
  scale_color_manual(values = c('turquoise', 'springgreen'))
```

The Nile Rivers Flow Over Time Daisy Cutter

