**Introduction to Tidyverse**

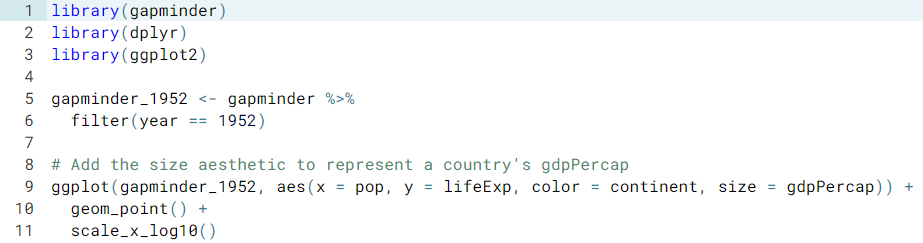
library (gplyr, gapminder, ggplot)

filter (year == 1987, country == "China")

arrange (Age) or arrange (desc(Age)) //sorts of ascending or descending order

% > % takes ourput from first command and gives it to the second

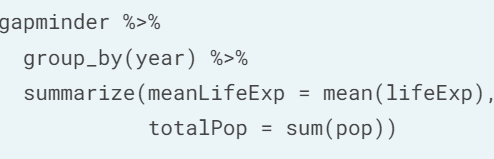
mutate (gdp = gdpPercap \* pop) //manipulate existing dataframes, create new colums



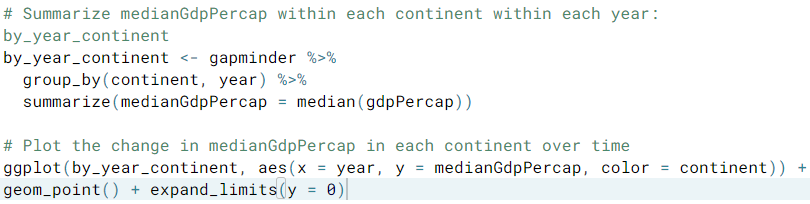
+ expand\_limits(y = 0) // y graph starts at 0

facet\_wrap(~ continent) // divides data on sub

summarize( meanLifeExp = mean(lifeExp), totalPop = sum(pop)) //mean – average

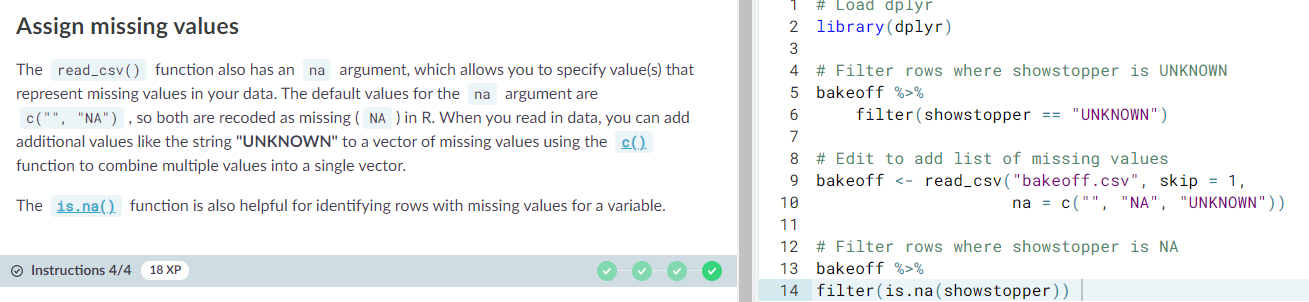
//median – point in a set of numbers, half are above and half below

group\_by(year, age) %>% //tells to do within each specified

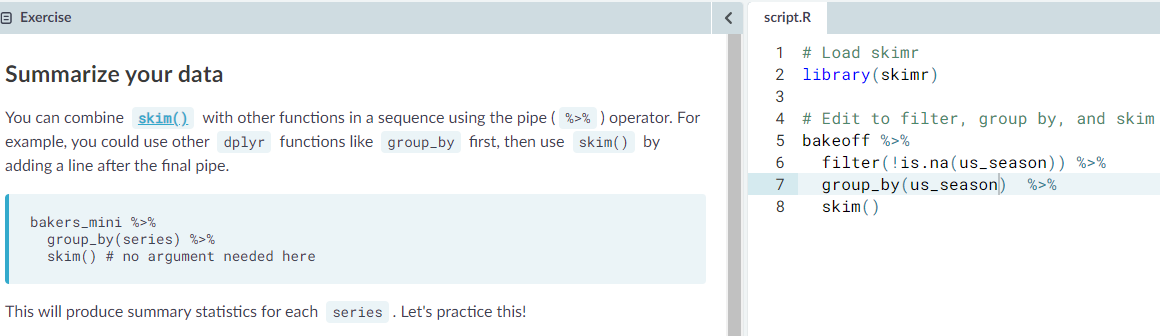


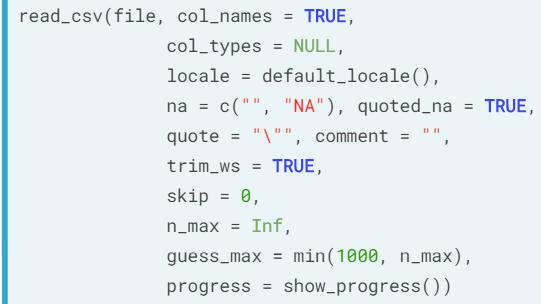
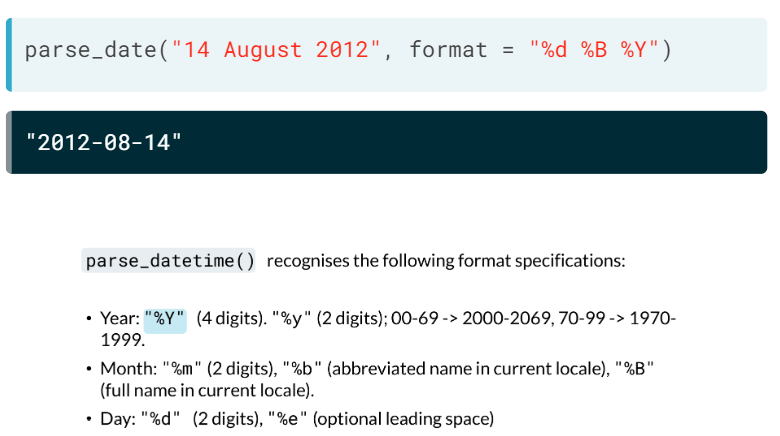
 geom\_col, geom\_histogram, geom\_line,

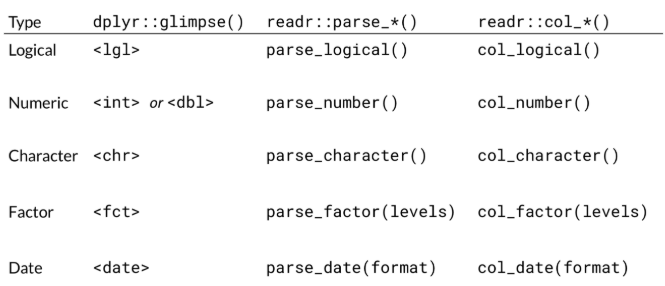
<https://ggplot2.tidyverse.org/reference/labs.html> - title and etc ,.

**Working with Data in Tidyverse**

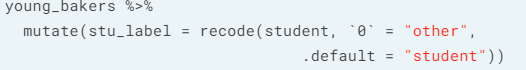
Glimpse () – show information in rows instead of columns

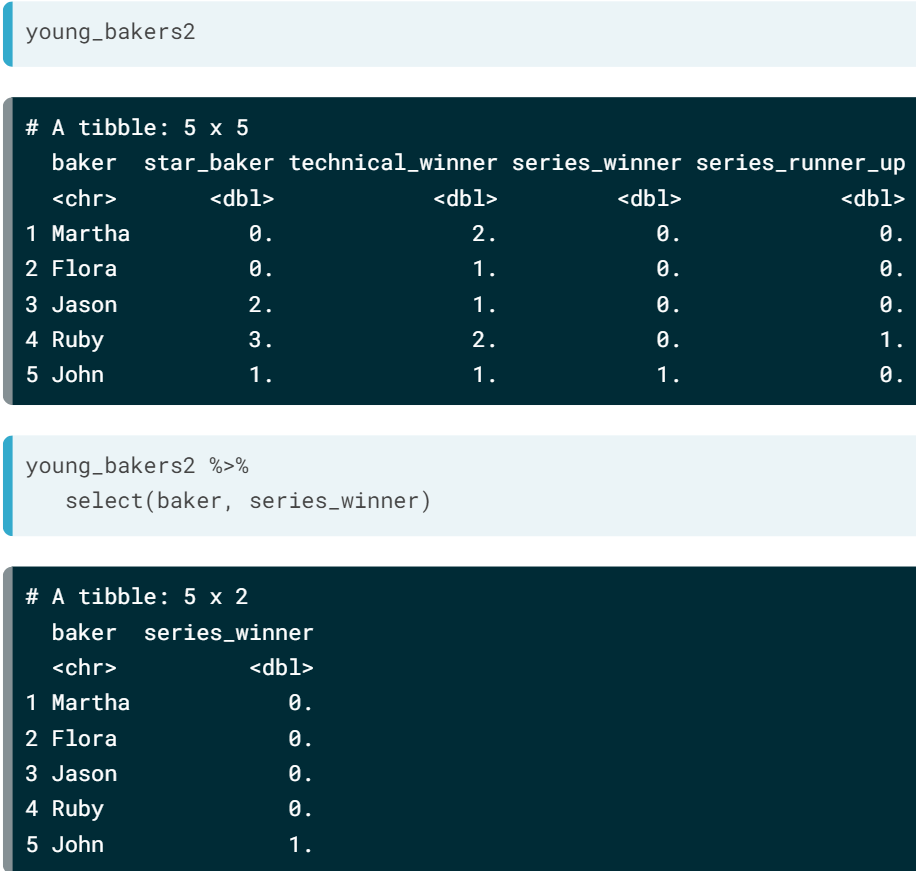
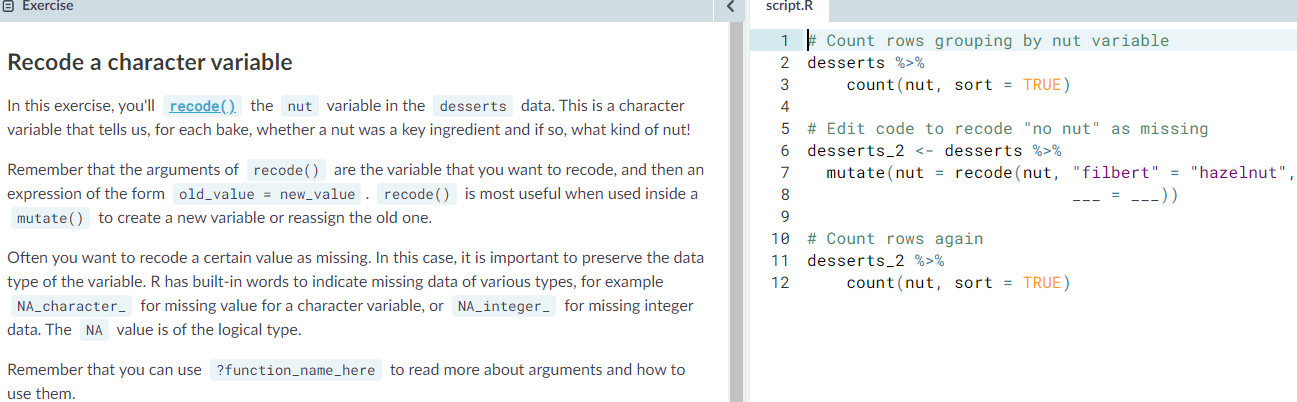
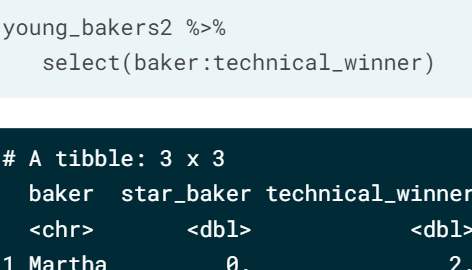
Bakeoff %>% Distinct (series)

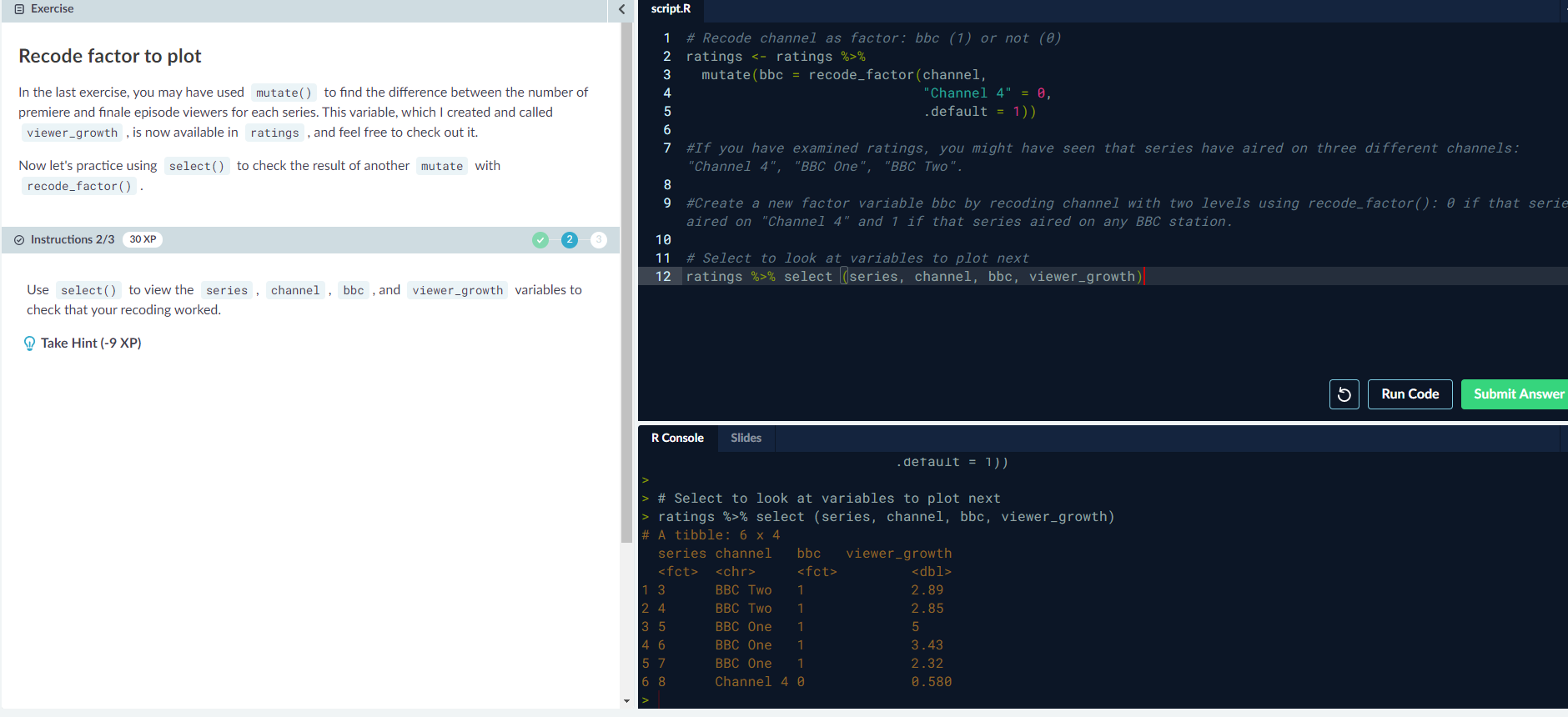
Count(object) – adds new column to count / parse

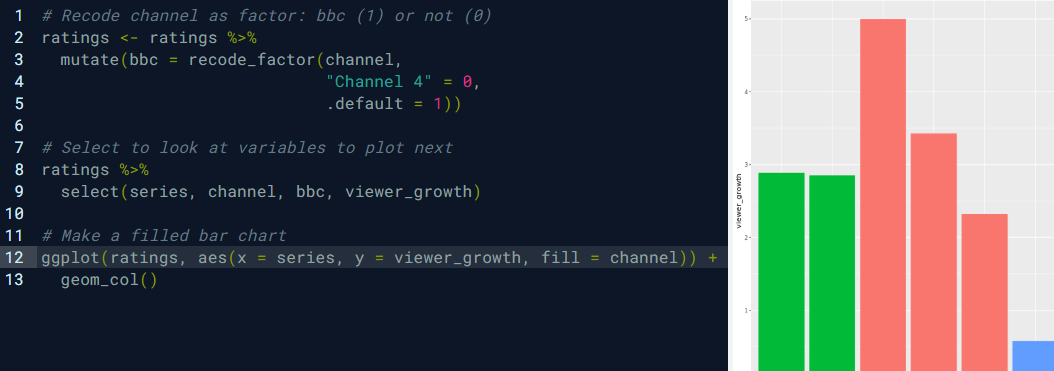
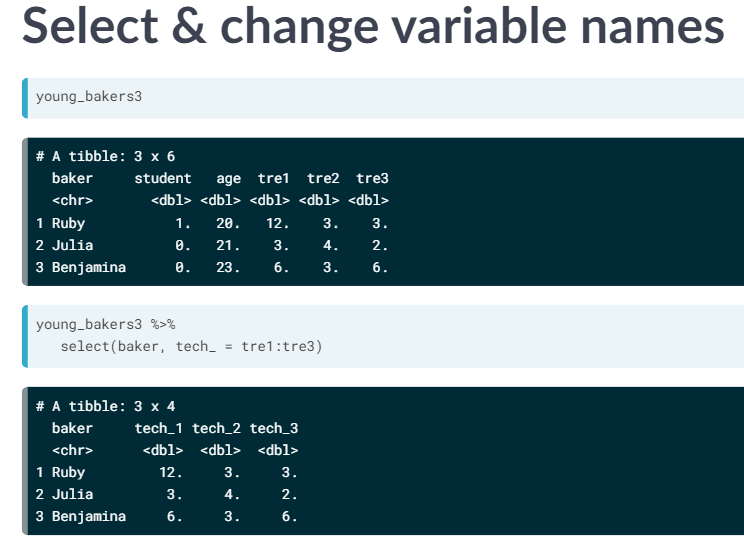
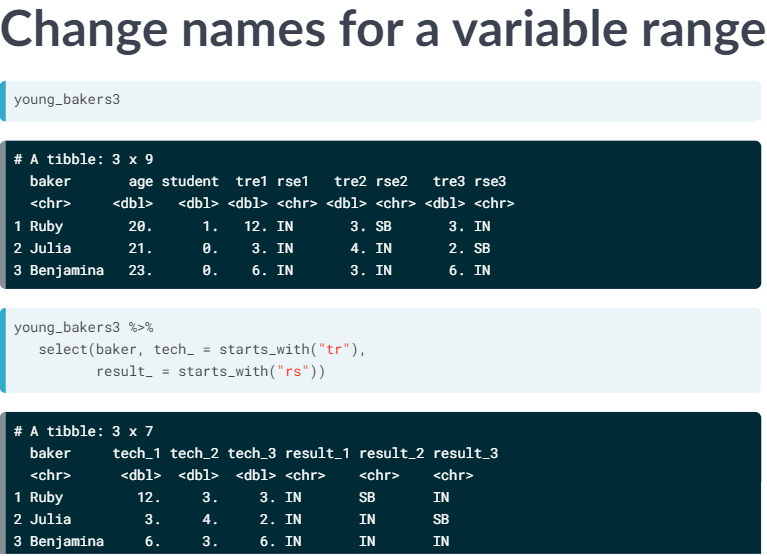
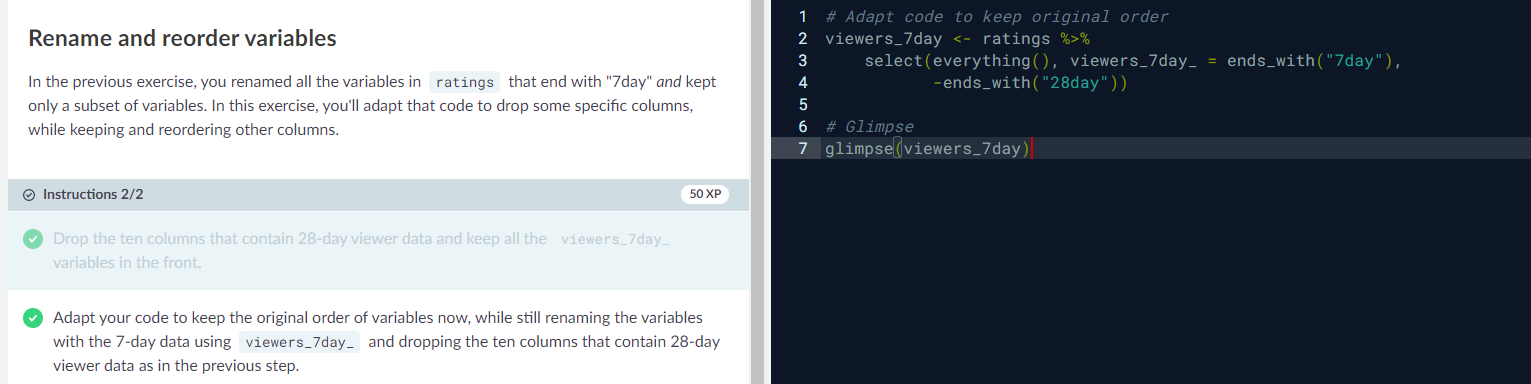
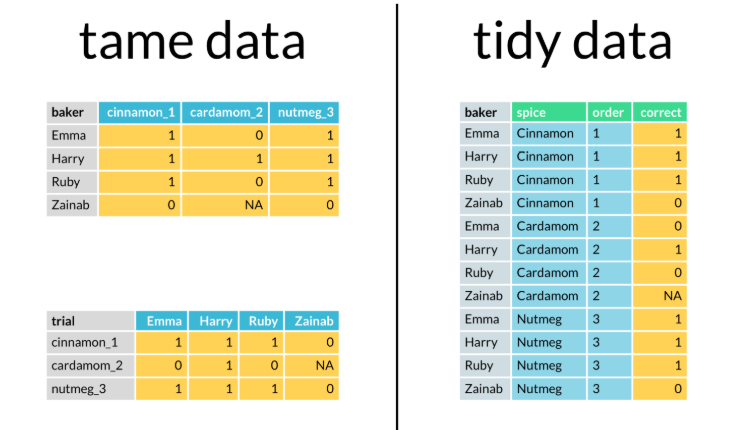


Cast?? Col?

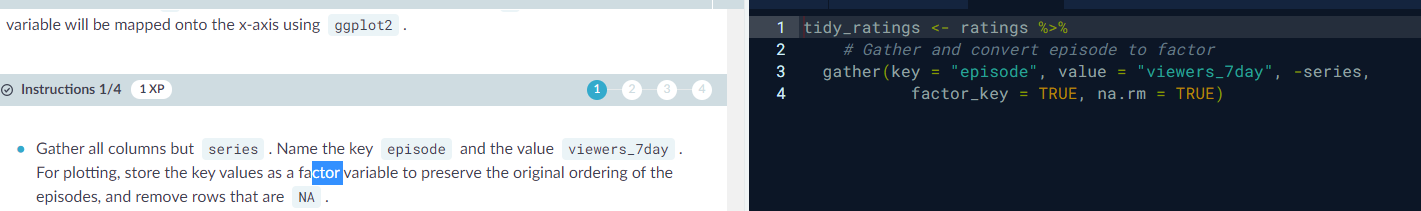


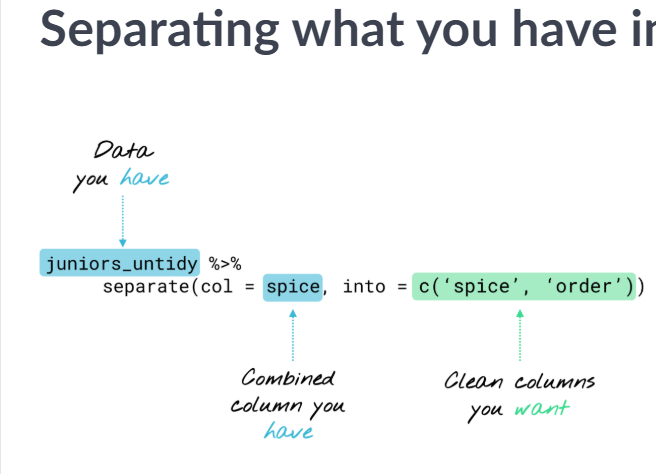
 select(-object) shows all except the –

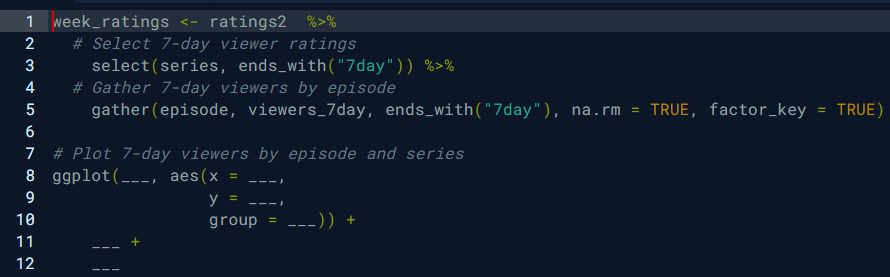
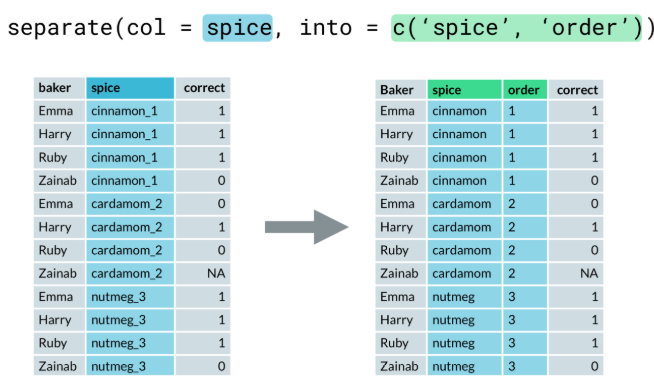
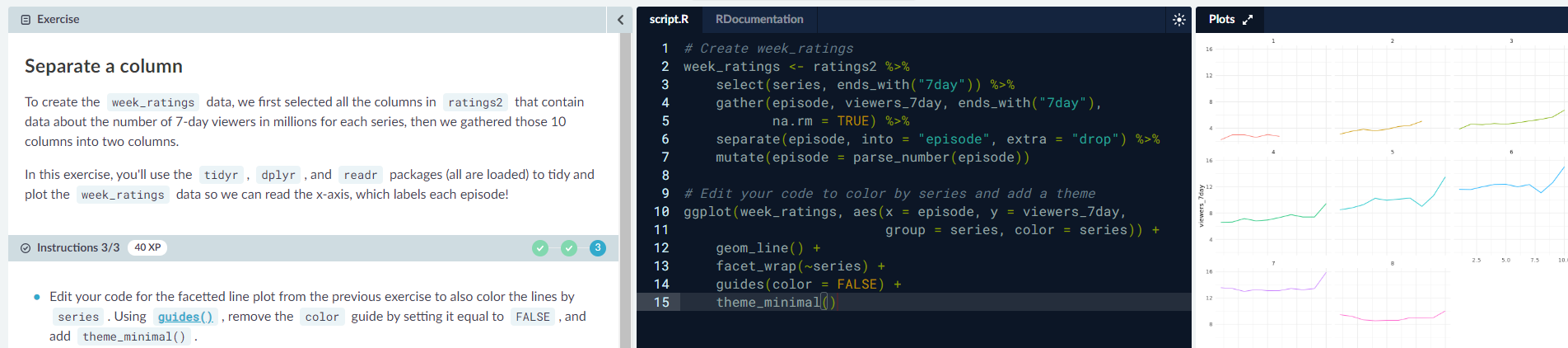
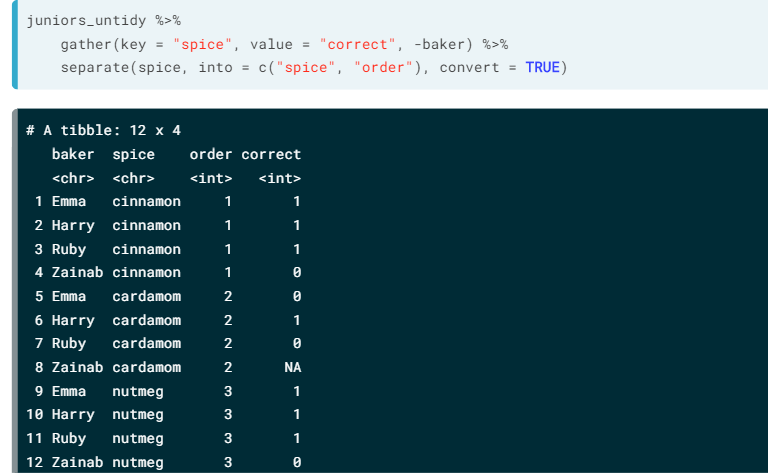
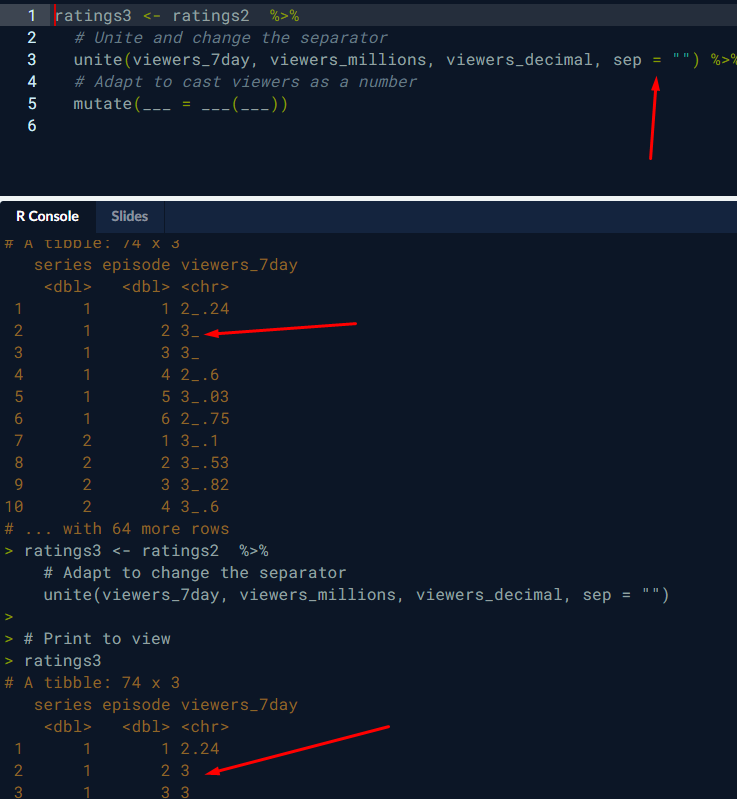
Select(baker, starts\_with(“series”)) // ends\_with() //contains()

  
rename(tech\_1 = t\_first)

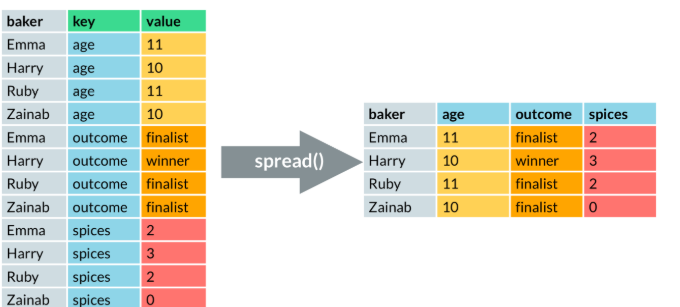
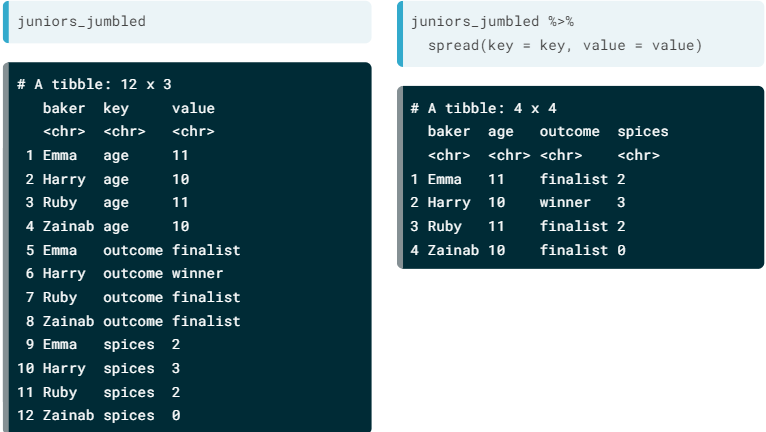


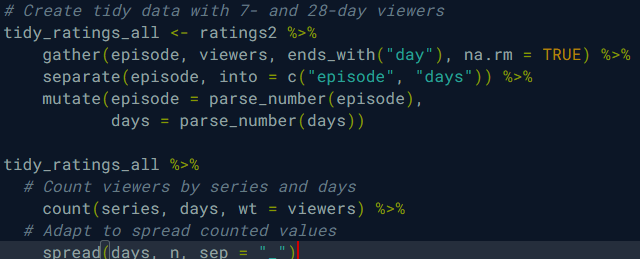
gather()

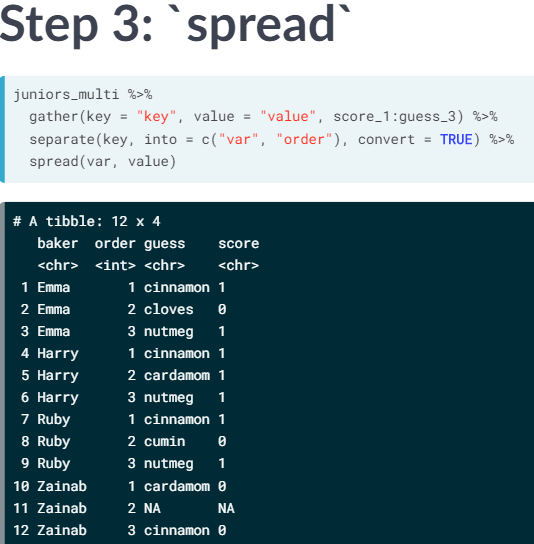
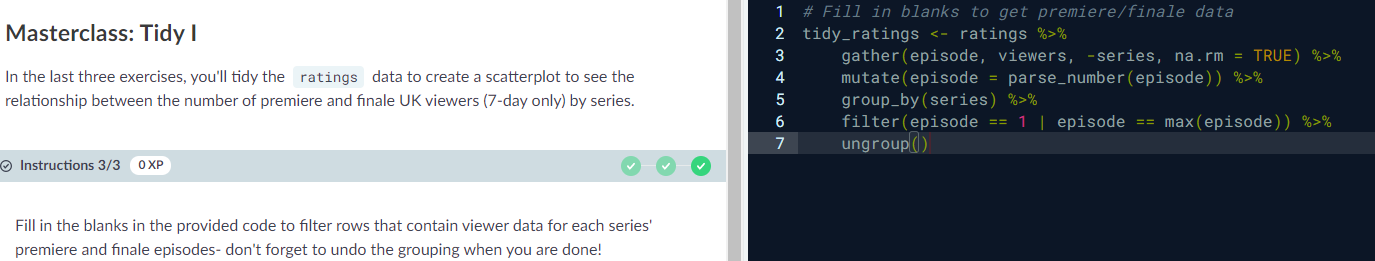
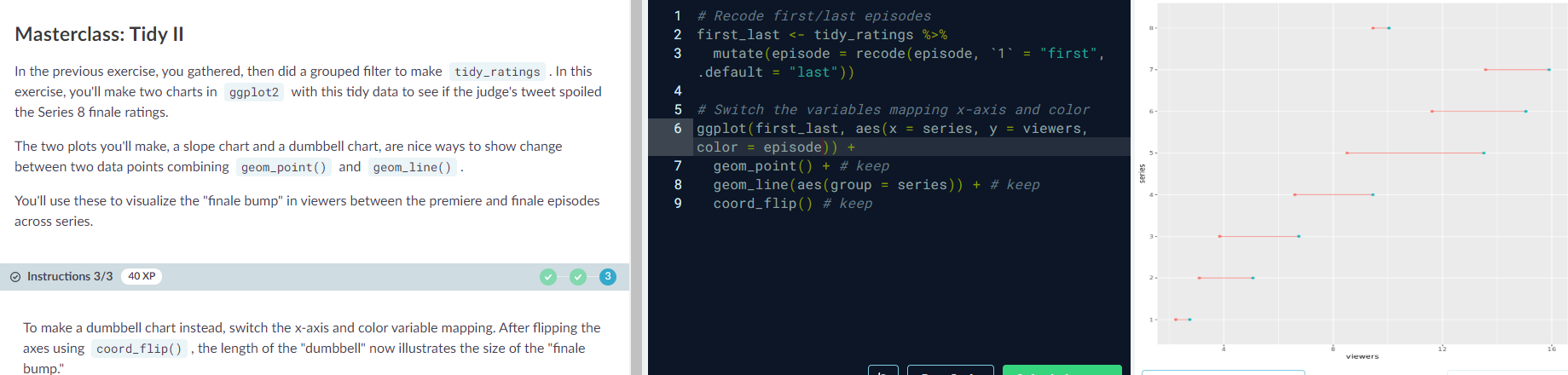


Row\_number / separate/ guide /

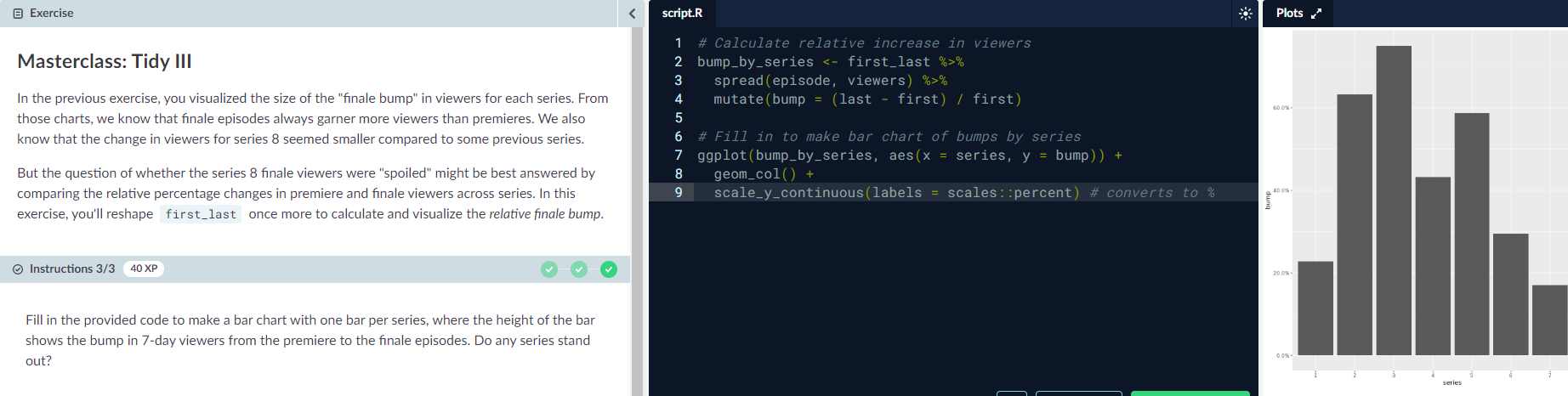
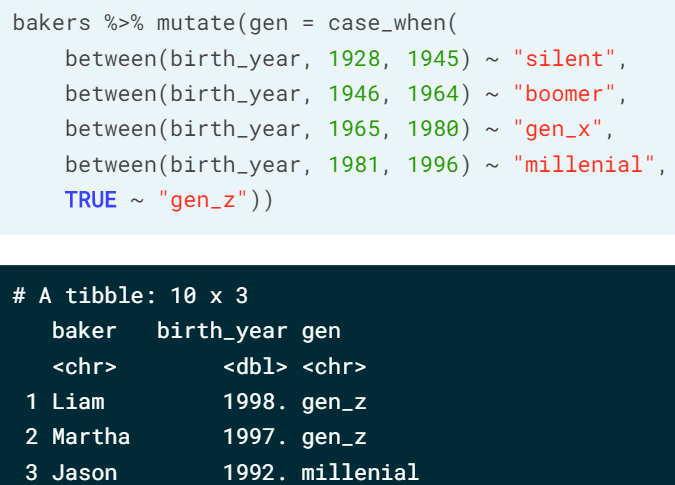
Unite / sep

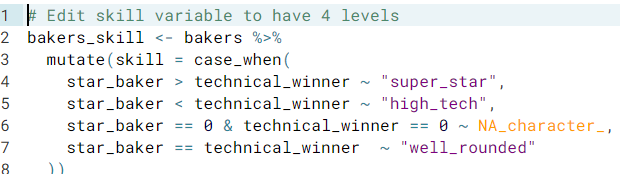
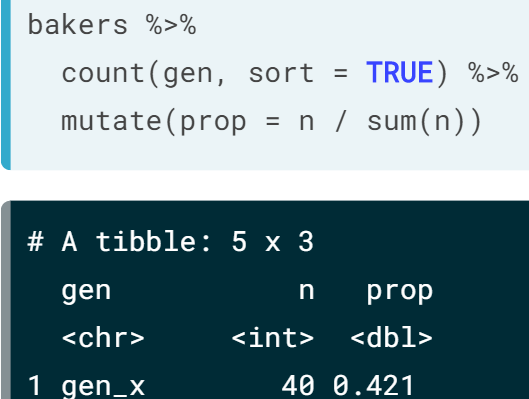
Opposide of Gather is Spread

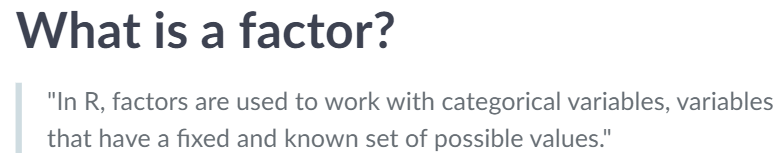
REMEMBER TO USE CONVERT INT -> NUM



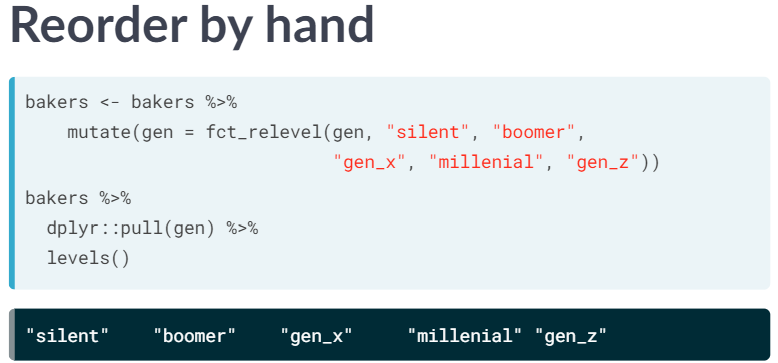


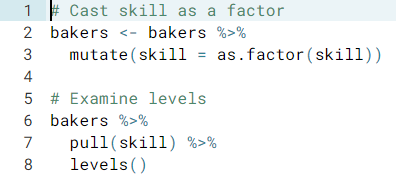
if else – case\_when

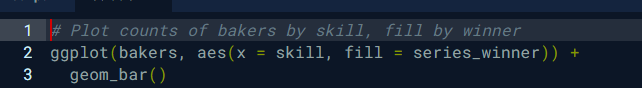
TRUE ~ “gen\_z” everything that is not in the mentioned ranges

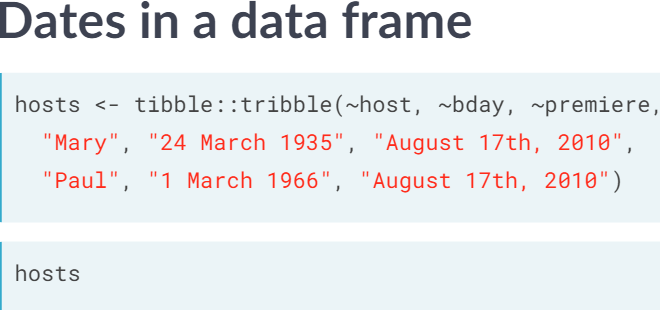
Factor

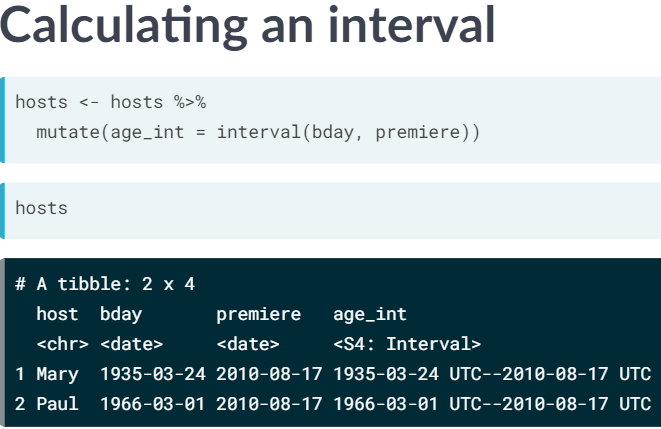
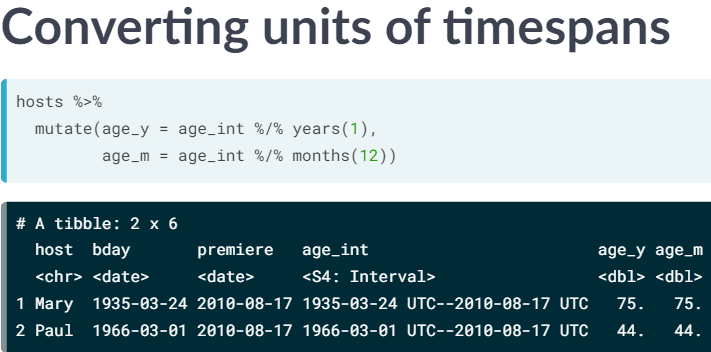
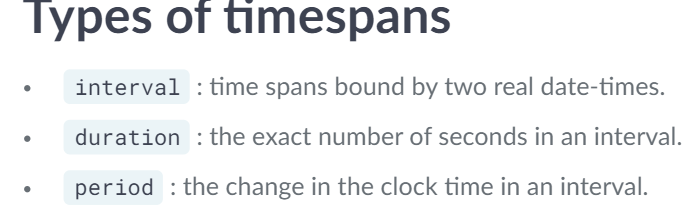
Levels only works with factor variables

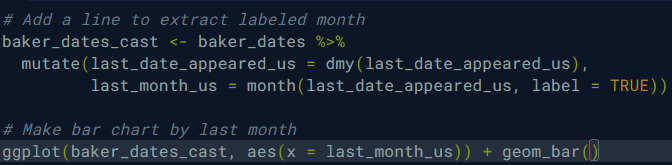


pull

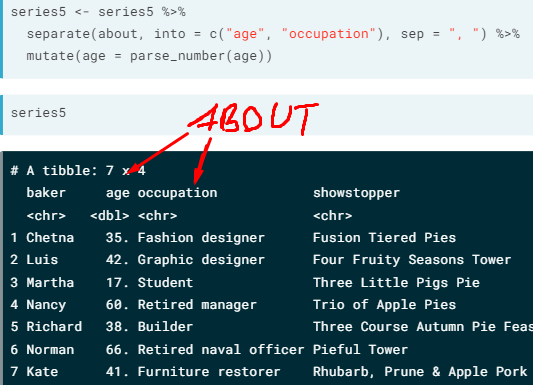


Dates

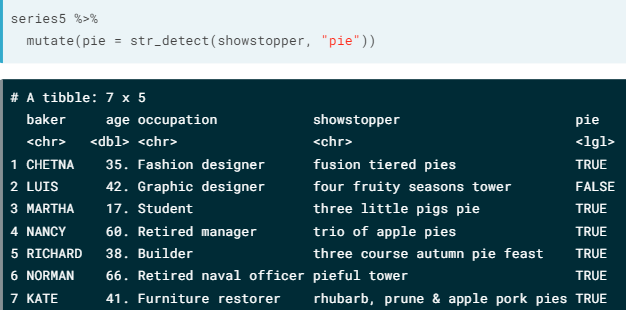
Calculate difference between two days – lubridate is the library

Month()



String/ Separate

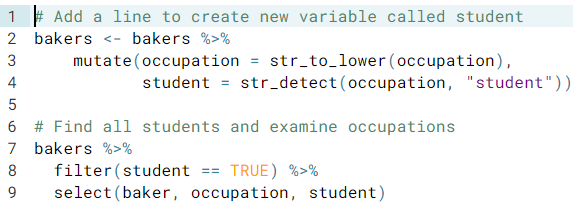
You can also use mutate(ear = str\_to\_upper(ear))

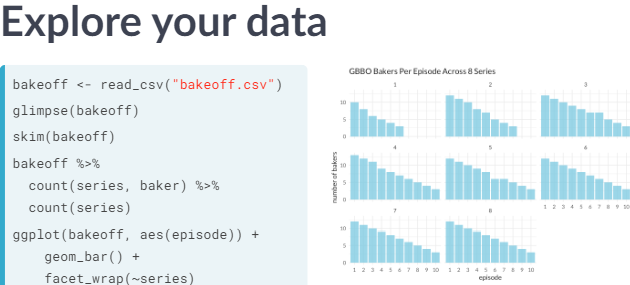
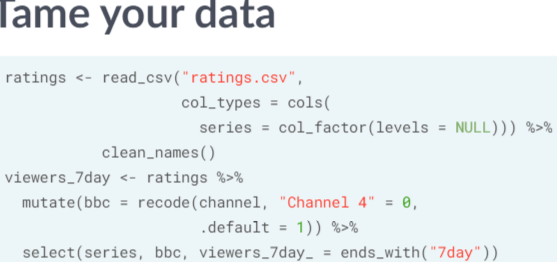
To make everything upper/lower case

You can use str\_detect

To check where is a word and always return logical . str\_trim(ear) to remove whitespace if have “”

There is also str\_replace(showstopper, “pie”, “cake”) and str\_remove()



  
Summary:

Data Visualisation with ggplot 2

Fine-tune scatter

##### Str() - Compactly Display The Structure Of An Arbitrary R Object

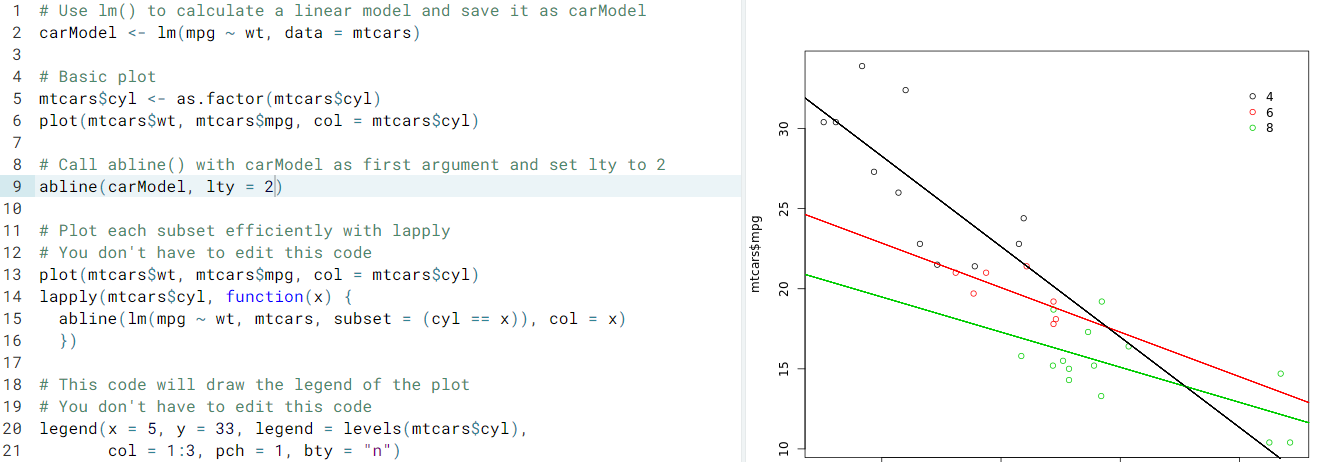
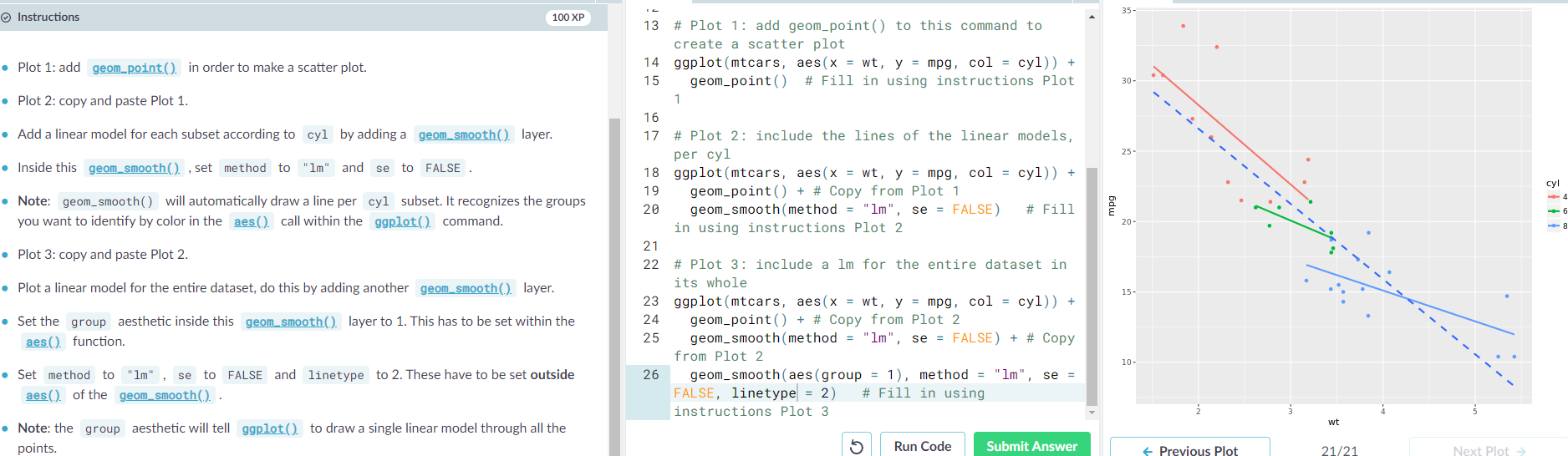
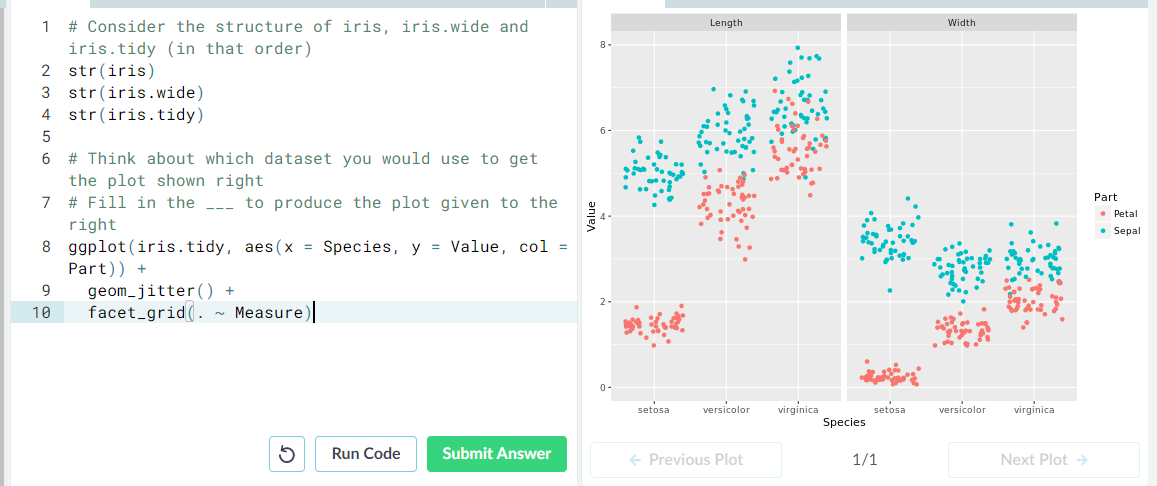
##### Grammatical Elements

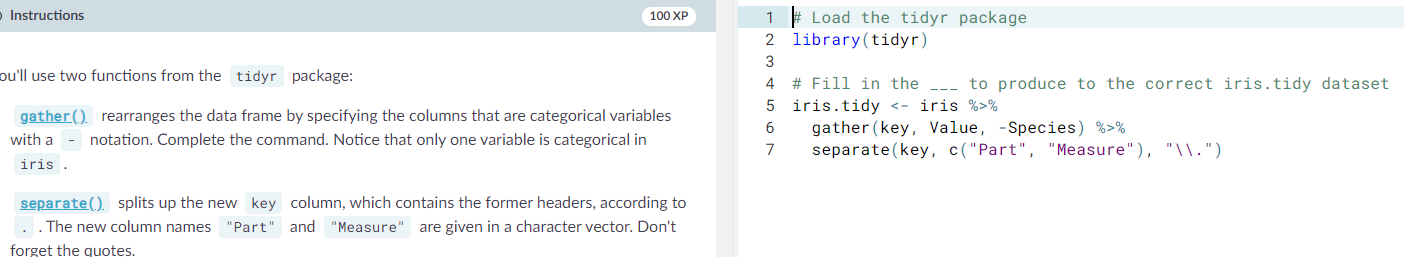
Geom\_smooth() – draws smooth lines over points

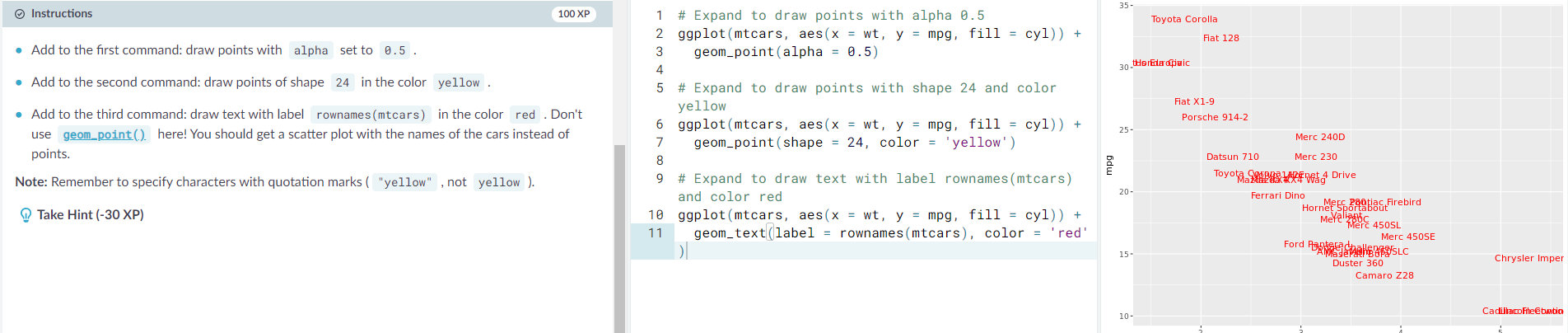


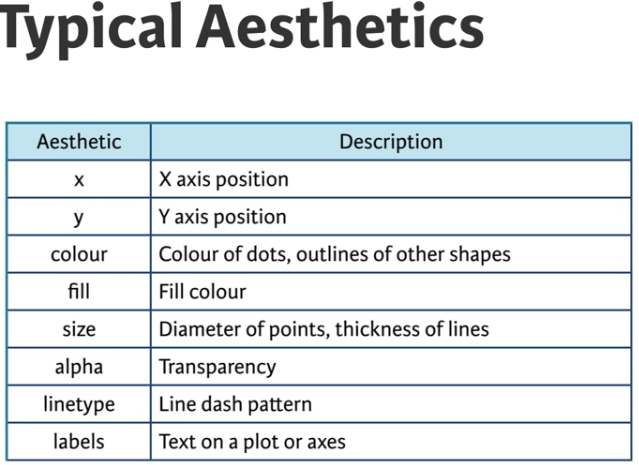
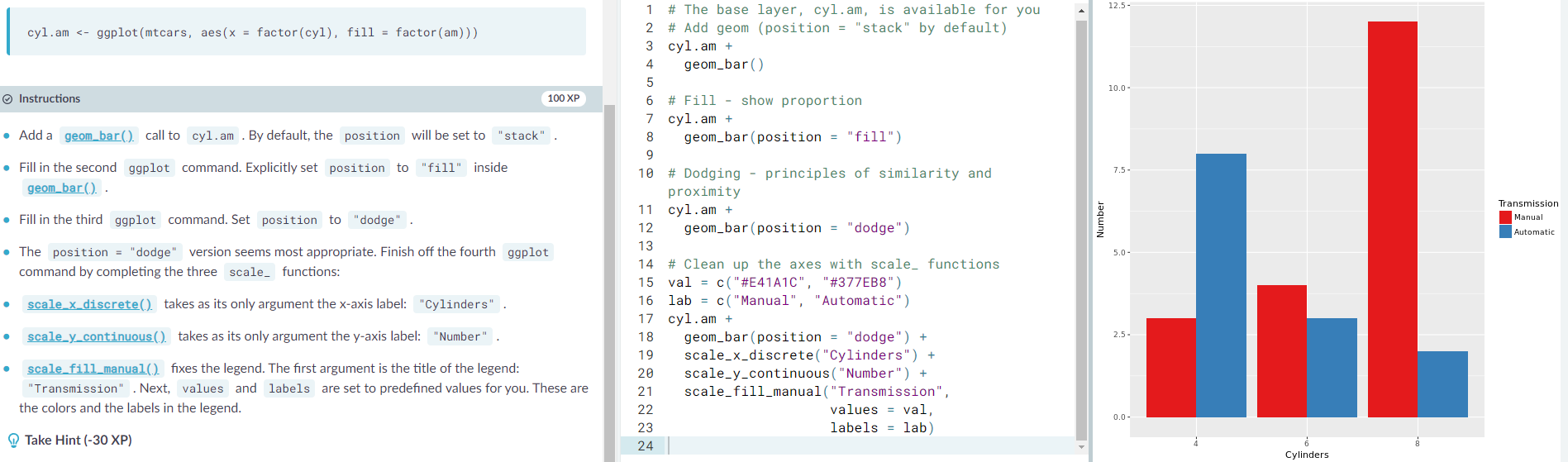
Base package – plot()

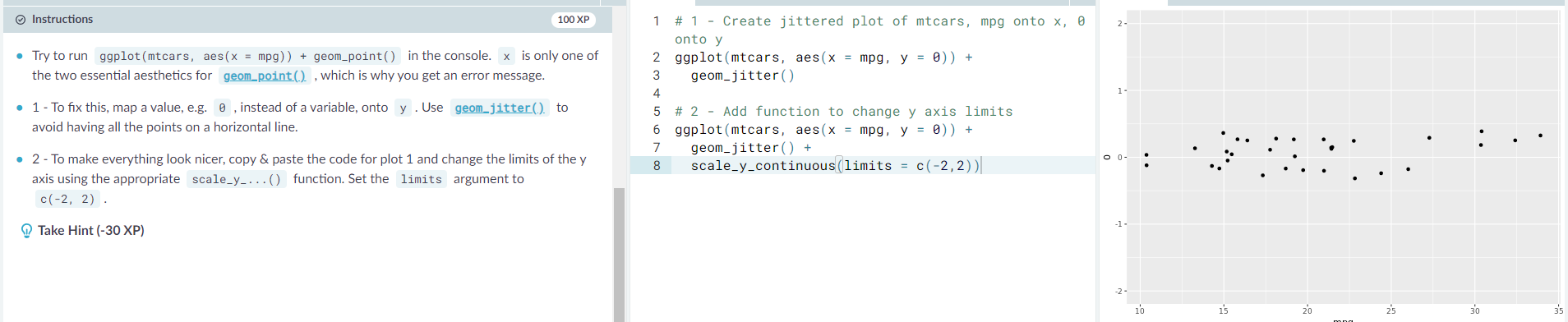
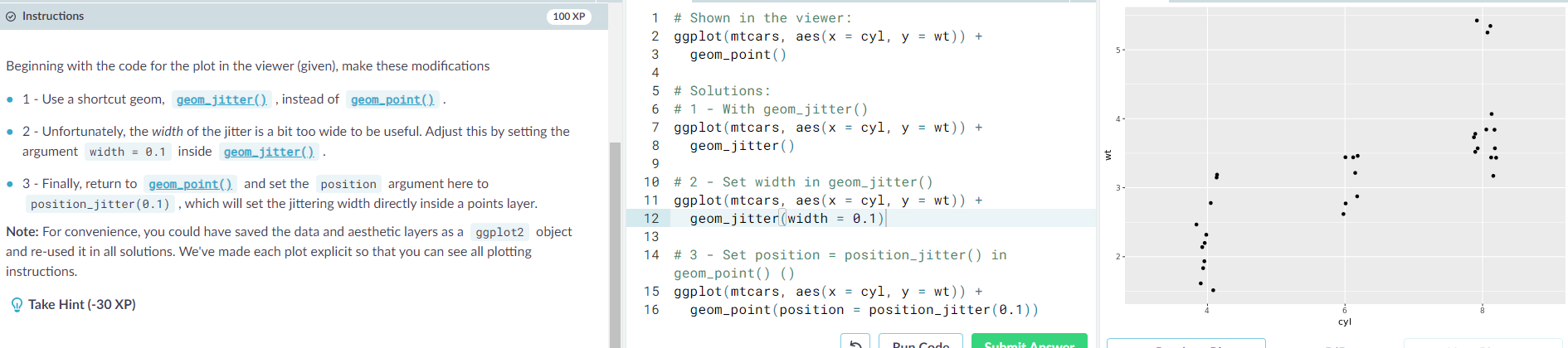
Abline() – adds straight line through current plot

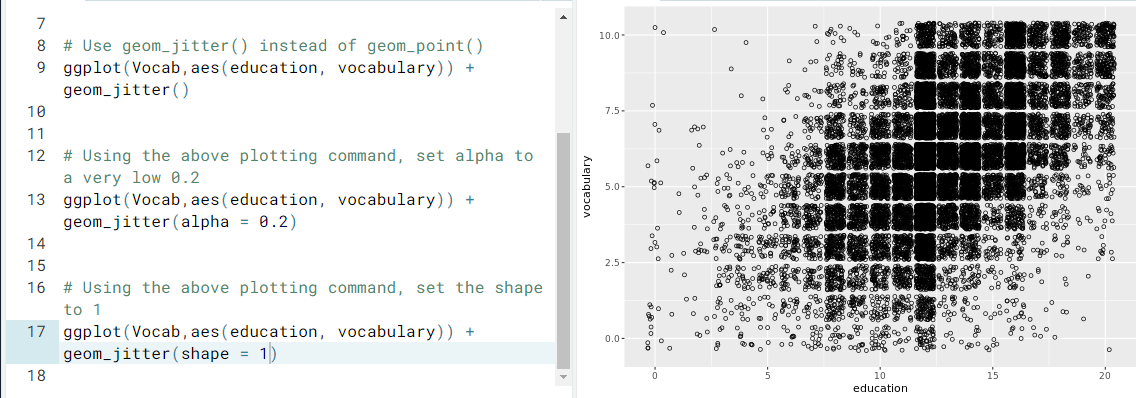




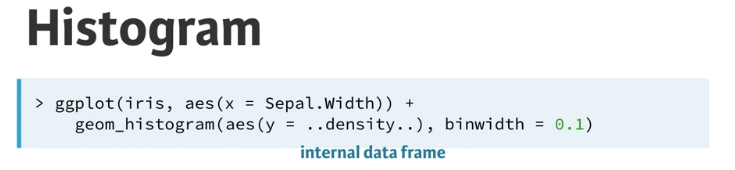
Aesthethics

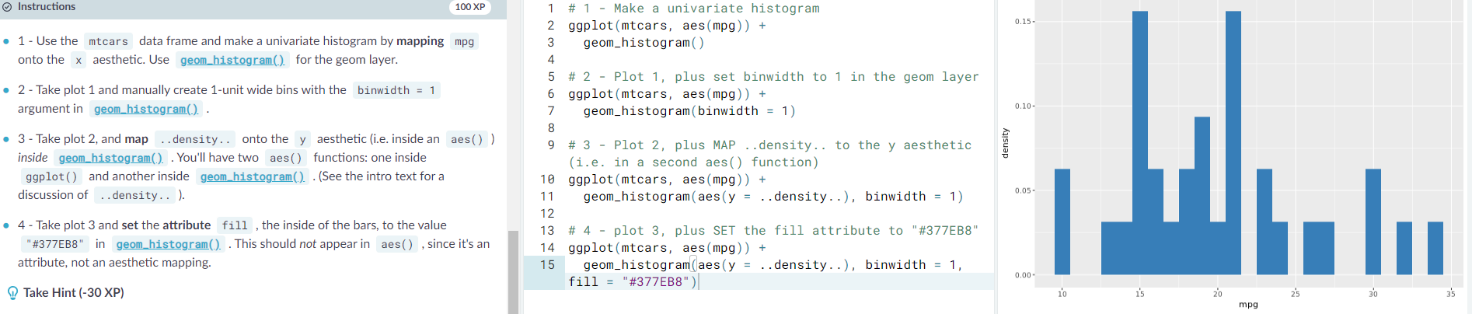
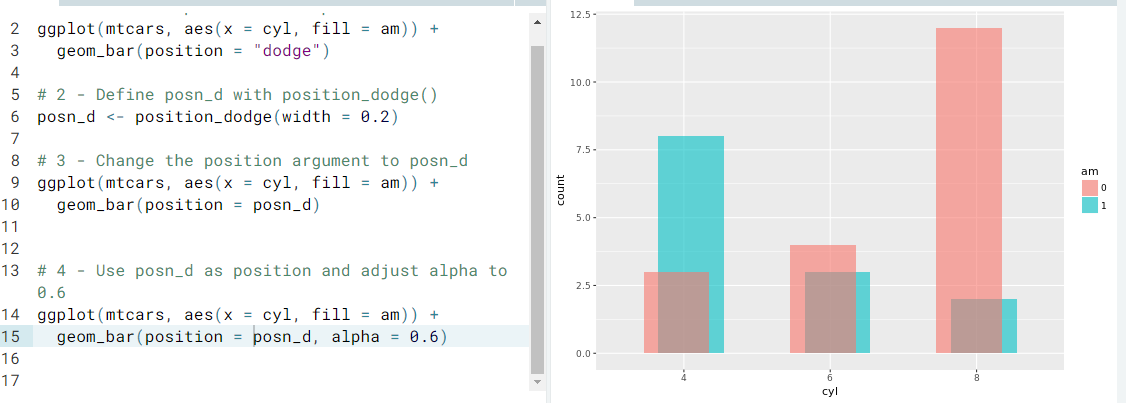
Scales

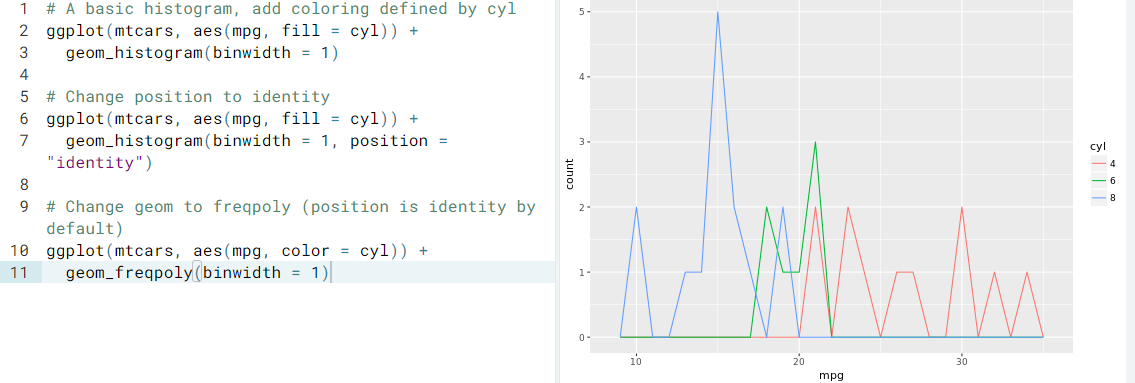
Axis limits, continuous // position\_jitter // dodge

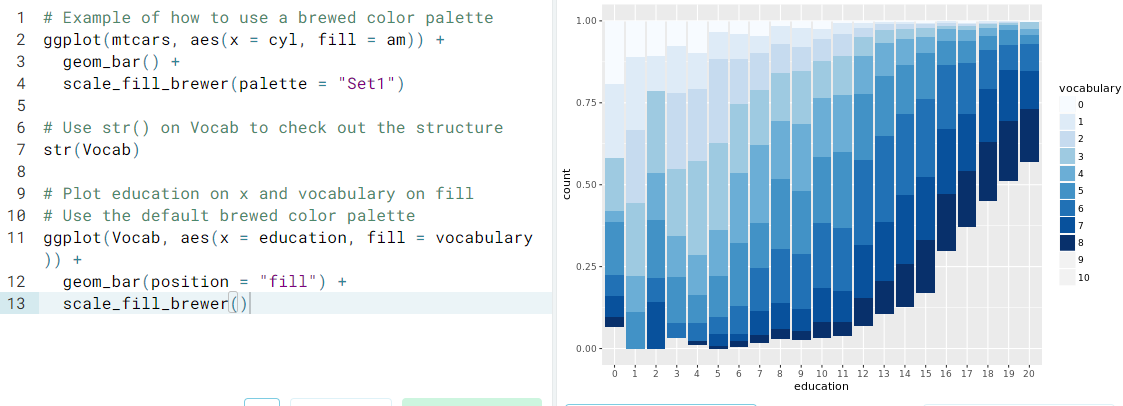


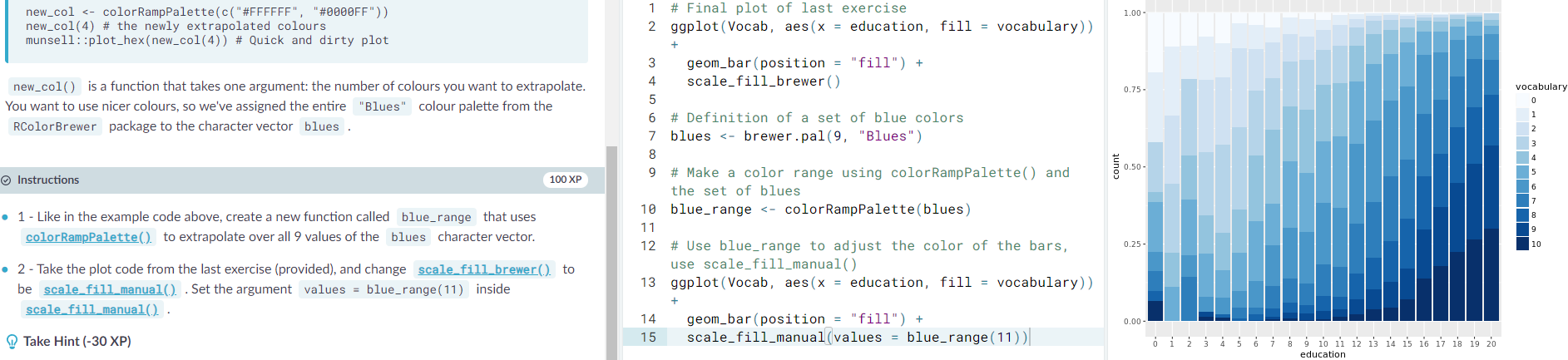
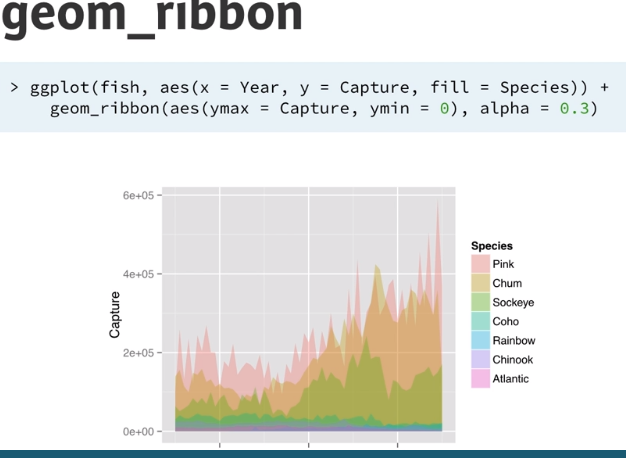
Histogram / binwidth



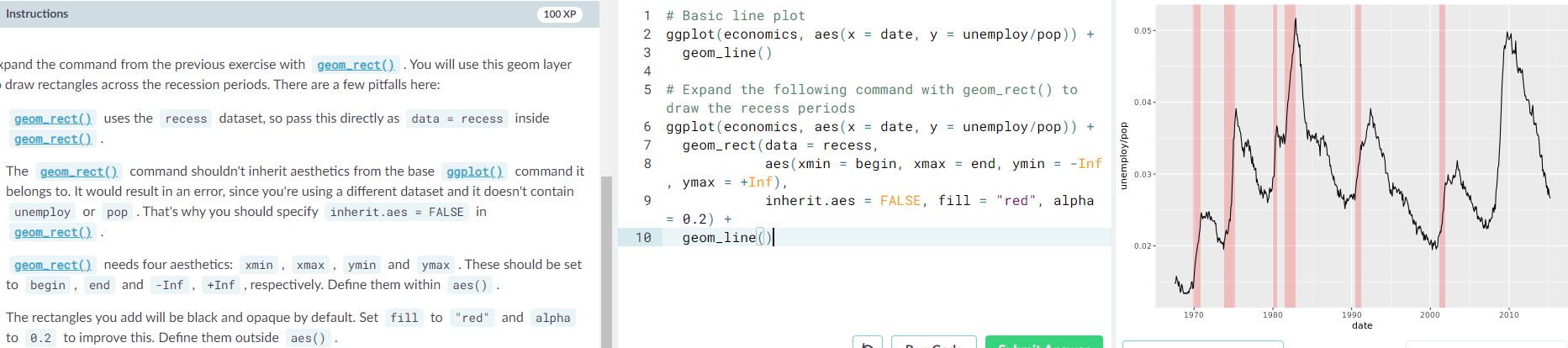
The ..density.. tells that the data is from the internal data frame

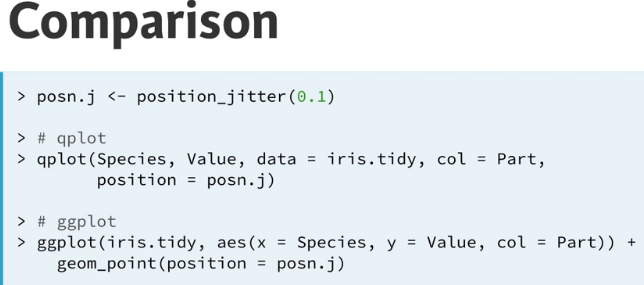






Geom\_rect()



Qplot()

