ECON03SEC1

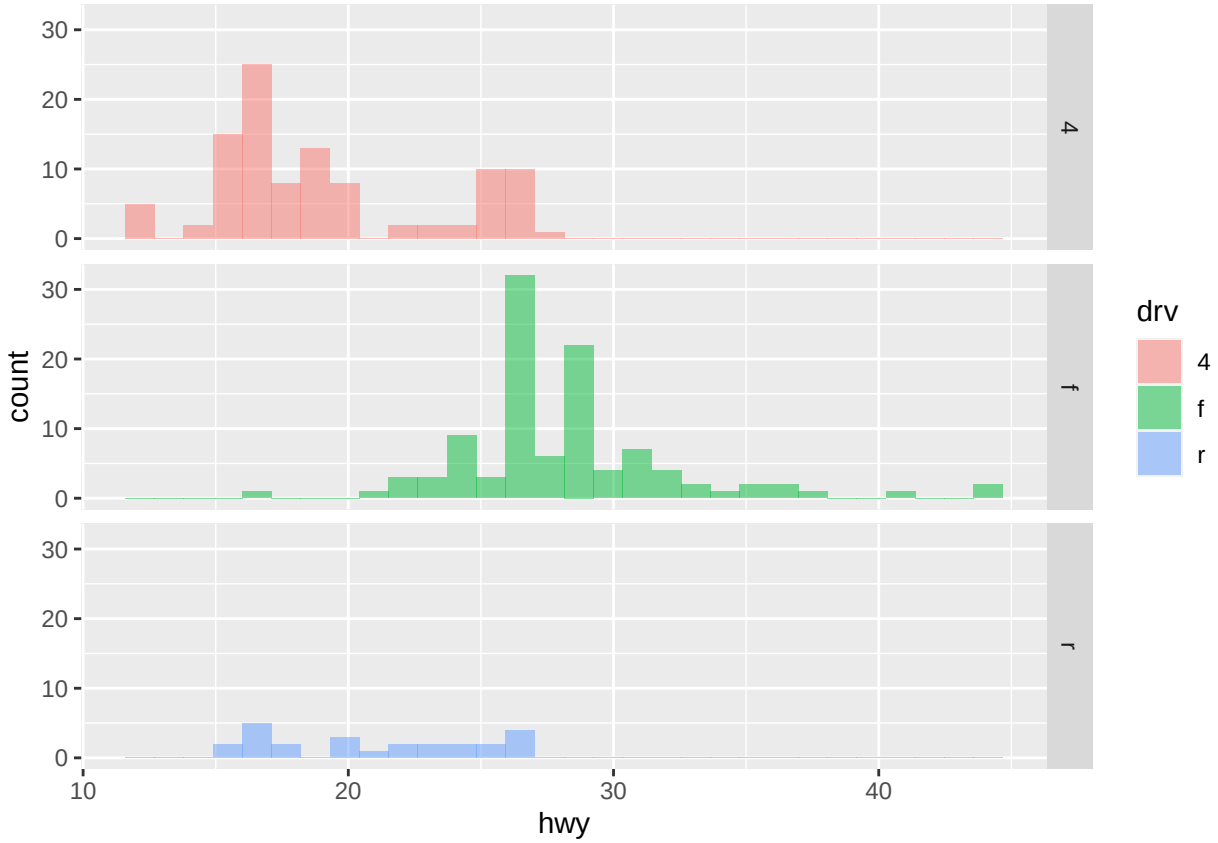
Internal Assessment 2

Full Marks: 50

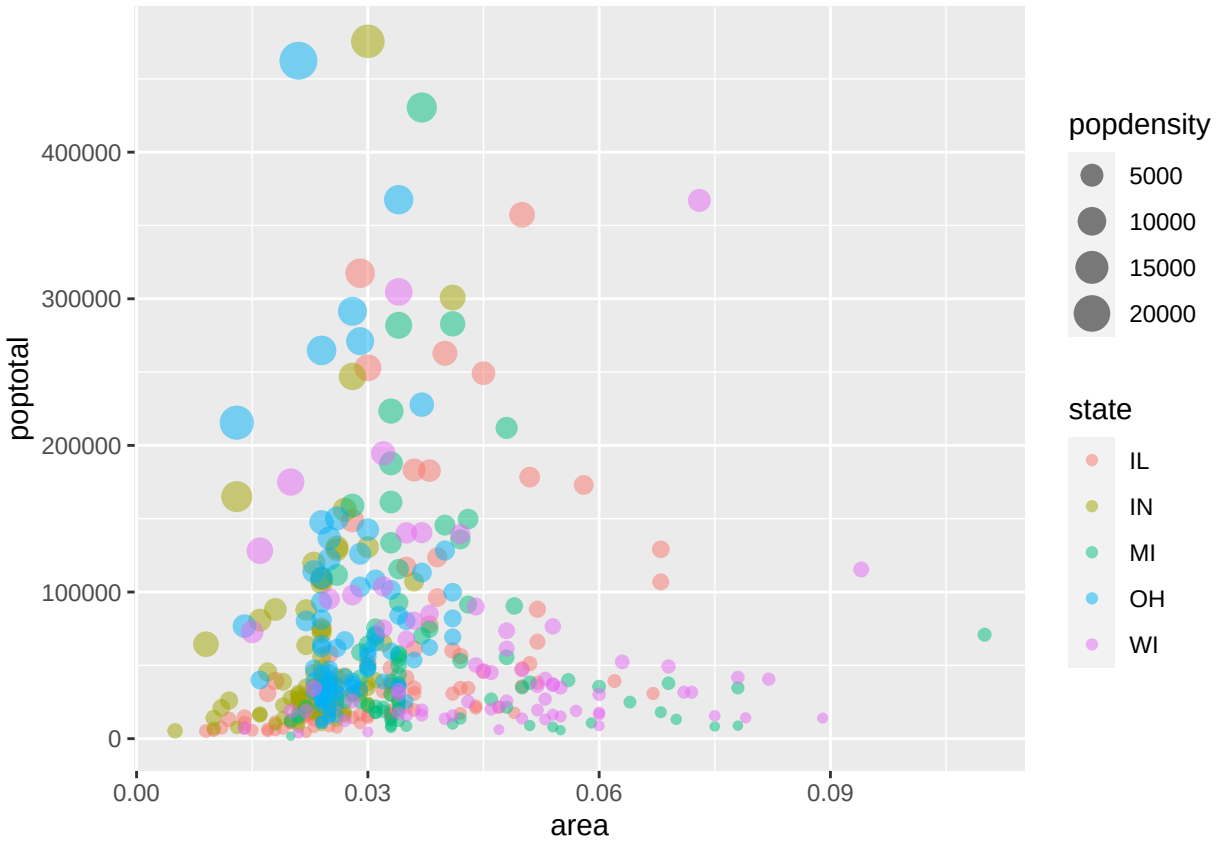
08/11/2023

`ggplot`

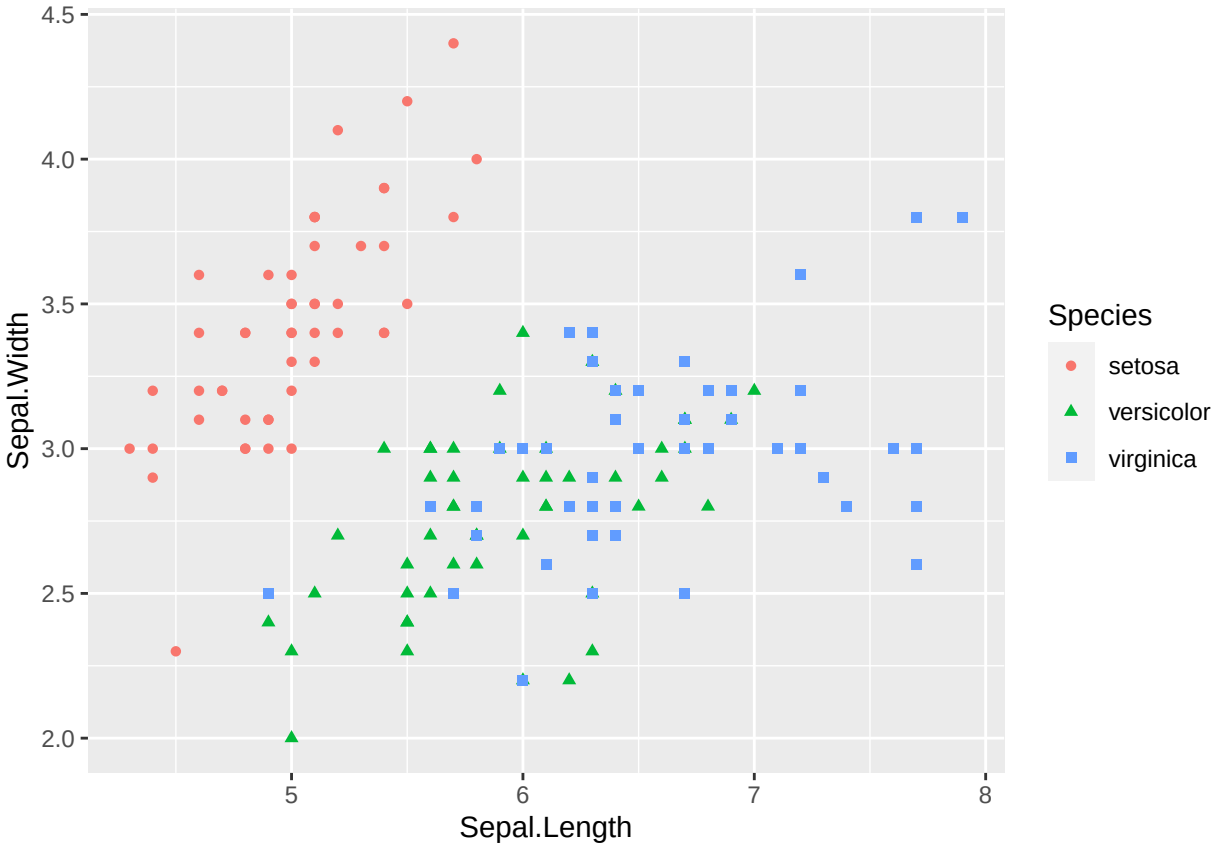
1. Using the `mpg`dataset in the `ggplot2` package, replicate the following plot.



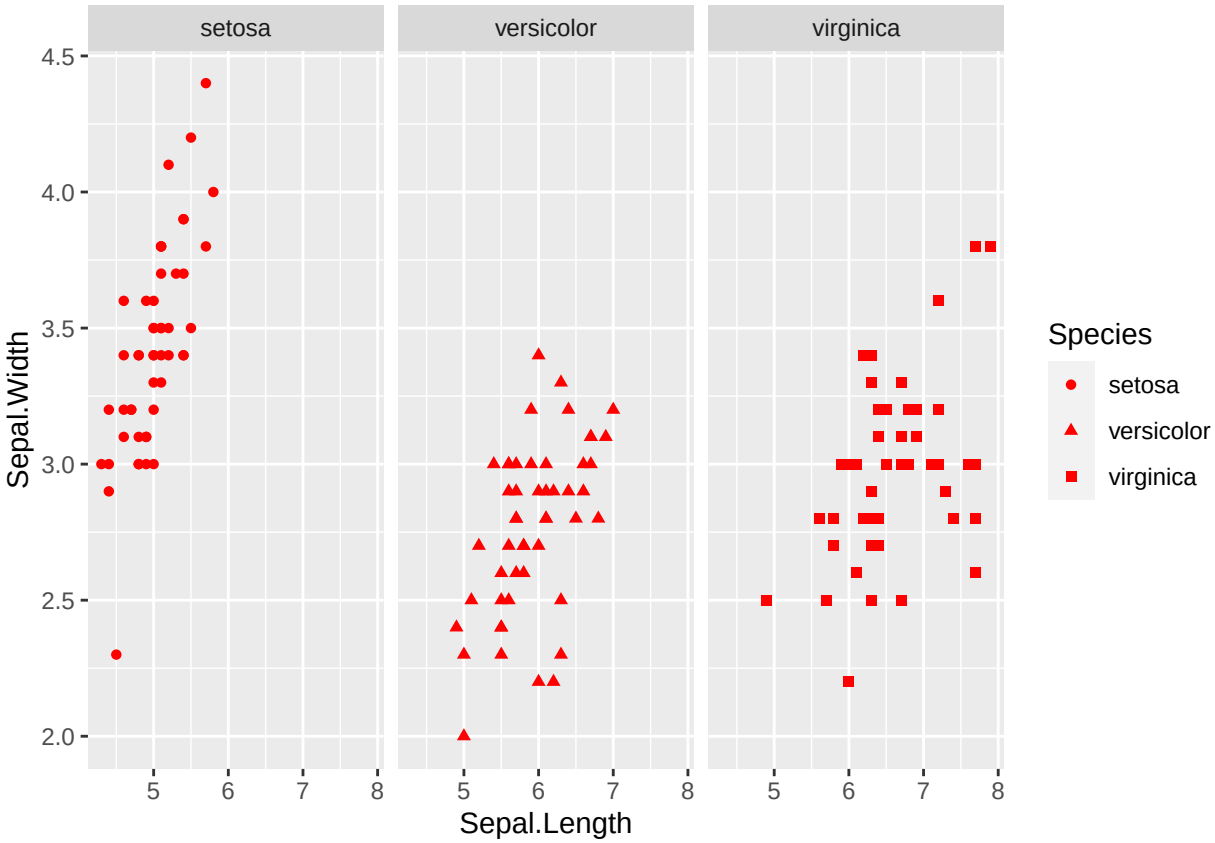
2. Using the `midwest` dataset in the `ggplot2` package, replicate the following plot.



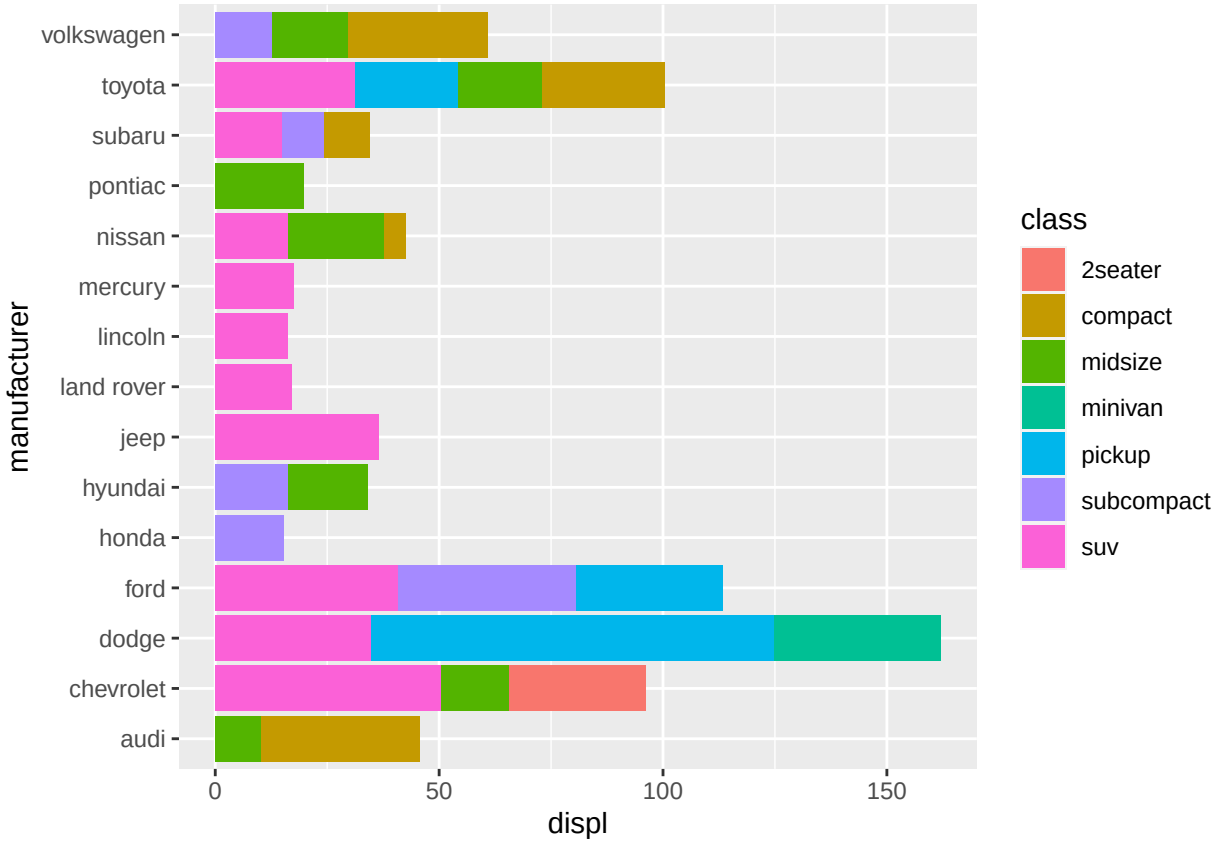
3. Using the `iris` dataset in the base R `datasets` package, replicate the following plot.



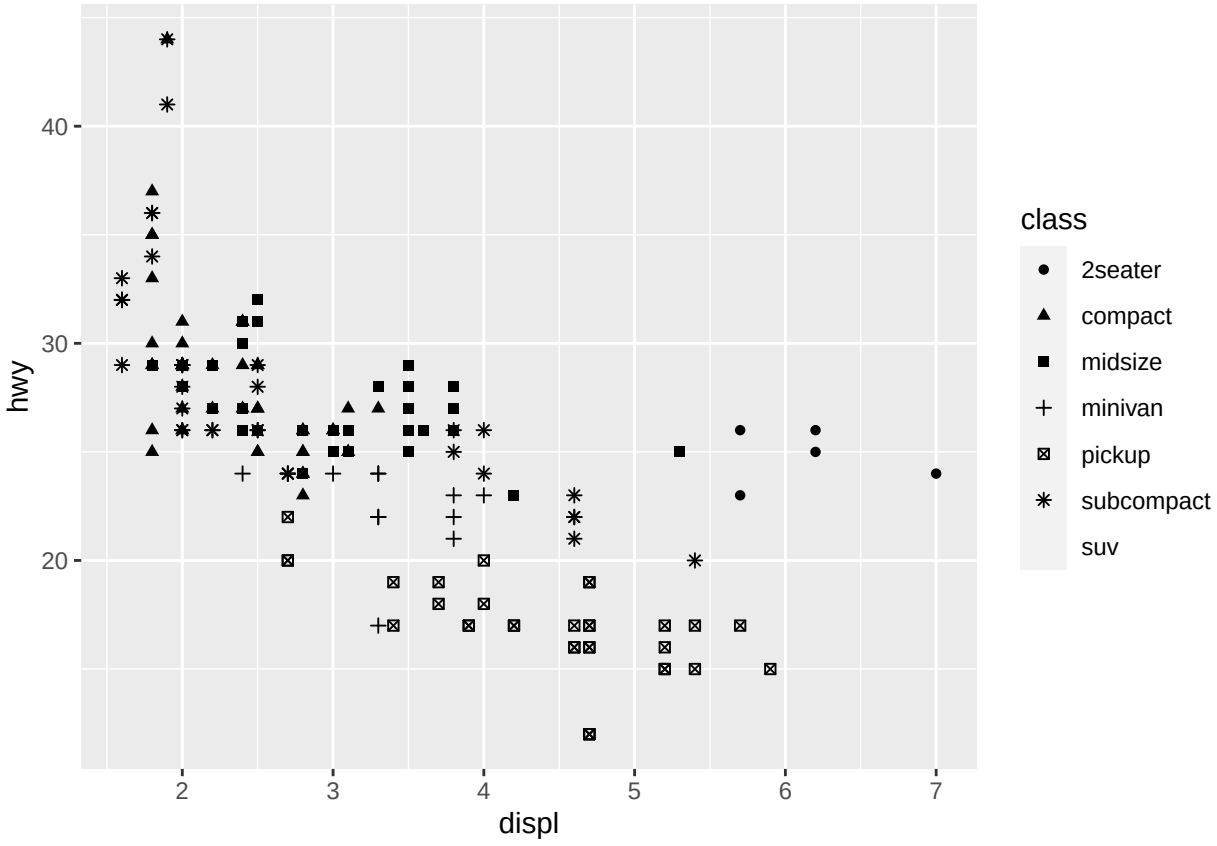
4. Using the `iris` dataset in the base R `datasets` package, replicate the following plot.



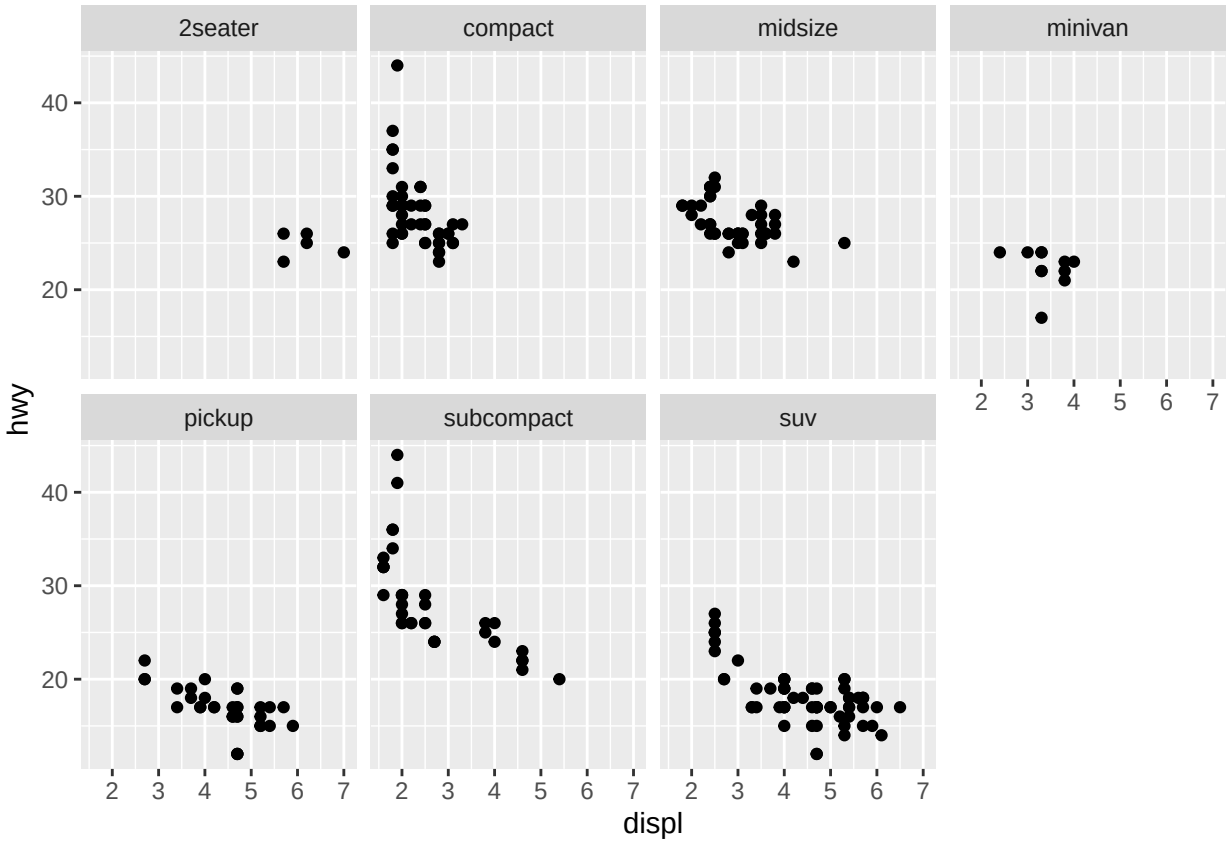
5. Using the `mpg` dataset in the `ggplot2` package, replicate the following plot.



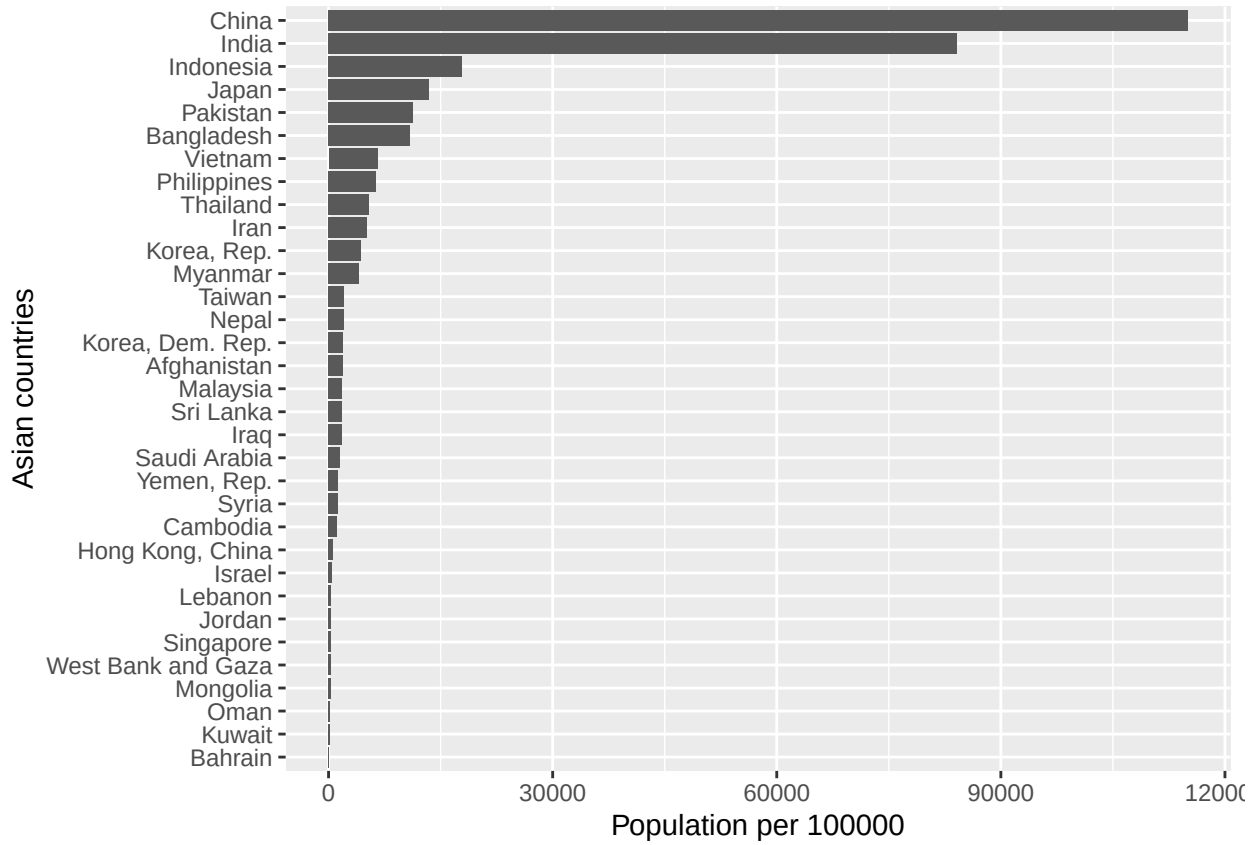
6. Using the `mpg` dataset in the `ggplot2` package, replicate the following plot.



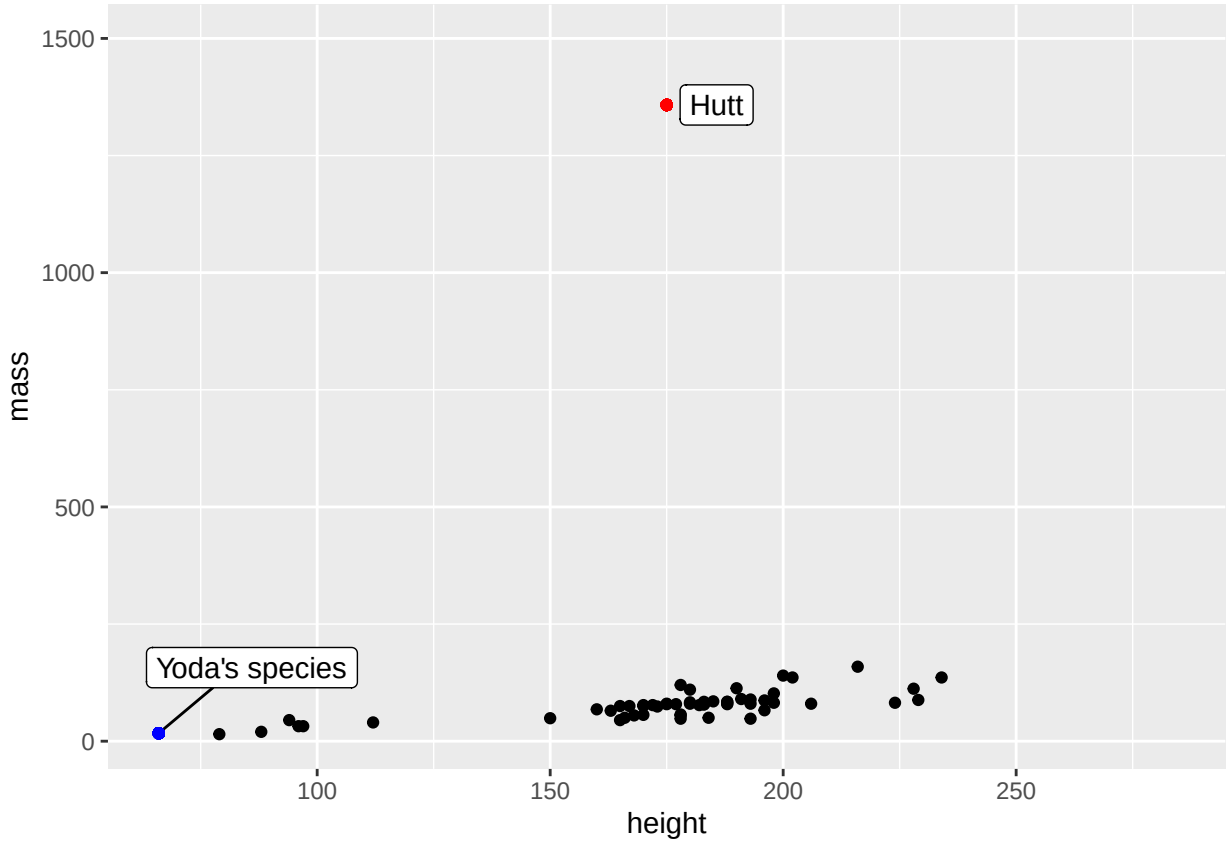
7. Using the `mpg` dataset in the `ggplot2` package, replicate the following plot.



8. Using the `gapminder` data in the `gapminder` package, replicate the following plot.



9. Using the `starwars` data in the `dplyr` package, replicate the following plot.



`dplyr`

1. Which species have blue eyes in the `starwars` dataset in the `dplyr` package?
2. How many female humans are there in the `starwars` dataset in the `dplyr` package ?
3. In the `starwars` dataset in the `dplyr` package, which species has the most number of blue eye colour?
4. In the `starwars` dataset in the `dplyr` package, what is the average mass of female Human species ?
5. In the `starwars` dataset in the `dplyr` package, how many species have a fair skin colour?
6. Calculate the mean *mpg* (miles per gallon) of the cars with 6 and 4 cylinders in `mtcars` dataset in the base R `datasets` package.
7. Which type of transmission (manual or automatic) has a higher variation (standard deviation) of *mpg* in the `mtcars` dataset in the base R `datasets` package?
8. What is the average displacement of a manual car with 4 cylinders in the `mtcars` dataset in the base R `datasets` package?
9. Which specie has the longest and widest petal in the `iris` dataset in the base R `datasets` packages?

`tidyr`

1. Tidy and replicate the `construction` dataset in the `tidyr` package as given below.

```
## # A tibble: 108 x 6
##     Year Month   Region    Completed_Units_Region Size      Completed_Units_Size
##    <dbl> <chr>   <chr>                      <dbl> <chr>                    <dbl>
##  1  2018 January Northeast                    114 1 unit                     859
##  2  2018 January Northeast                    114 2 to 4 u~                   NA
##  3  2018 January Northeast                    114 5 units ~                  348
##  4  2018 January Midwest                      169 1 unit                     859
##  5  2018 January Midwest                      169 2 to 4 u~                   NA
##  6  2018 January Midwest                      169 5 units ~                  348
##  7  2018 January South                        596 1 unit                     859
##  8  2018 January South                        596 2 to 4 u~                   NA
##  9  2018 January South                        596 5 units ~                  348
## 10  2018 January West                         339 1 unit                     859
## # i 98 more rows
```

2. Tidy and replicate the `fish_encounters` dataset in the `tidyr` package as given below.

```
## # A tibble: 5 x 12
##   fish  Release I80_1 Lisbon  Rstr Base_TD   BCE   BCW  BCE2  BCW2   MAE   MAW
##   <fct>   <int> <int>  <int> <int>   <int> <int> <int> <int> <int> <int> <int>
## 1 4842        1     1      1     1       1     1     1     1     1     1     1
## 2 4843        1     1      1     1       1     1     1     1     1     1     1
## 3 4844        1     1      1     1       1     1     1     1     1     1     1
## 4 4858        1     1      1     1       1     1     1     1     1     1     1
## 5 4861        1     1      1     1       1     1     1     1     1     1     1
```

3. Tidy and replicate the `who` dataset in the `tidyr` package as given below.

```
## # A tibble: 56 x 6
##    country iso2  iso3   year Diagnosis    Value
##    <chr>   <chr> <chr> <dbl> <chr>        <dbl>
##  1 India   IN    IND    2002 new_sp_m3544 55829
##  2 India   IN    IND    2002 new_sp_m2534 54719
##  3 India   IN    IND    2002 new_sp_m4554 44532
##  4 India   IN    IND    2002 new_sp_m1524 39923
##  5 India   IN    IND    2002 new_sp_f2534 31946
##  6 India   IN    IND    2002 new_sp_f1524 28573
##  7 India   IN    IND    2002 new_sp_m5564 28199
##  8 India   IN    IND    2002 new_sp_f3544 21378
##  9 India   IN    IND    2002 new_sp_m65   14960
## 10 India   IN    IND    2002 new_sp_f4554 13233
## # i 46 more rows
```

4. Tidy and replicate the `world_bank_pop` dataset in the `tidyr` package as given below.

```
## # A tibble: 4 x 4
##   country indicator   Year     Population
##   <chr>   <chr>       <chr>         <dbl>
## 1 IND     SP.POP.GROW 2000           1.82
## 2 IND     SP.URB.GROW 2000           2.60
## 3 IND     SP.URB.TOTL 2000   293168849
## 4 IND     SP.POP.TOTL 2000  1059633675
```

5. Tidy and replicate the `population` dataset in the `tidyr` package as given below.

```
## # A tibble: 2 x 20
##   country  `1995` `1996` `1997` `1998` `1999` `2000` `2001` `2002` `2003` `2004`
##   <chr>     <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>
## 1 India    9.56e8 9.73e8 9.90e8 1.01e9 1.03e9 1.04e9 1.06e9 1.08e9 1.09e9 1.11e9
## 2 China    1.24e9 1.25e9 1.26e9 1.27e9 1.27e9 1.28e9 1.29e9 1.30e9 1.30e9 1.31e9
## # i 9 more variables: `2005` <dbl>, `2006` <dbl>, `2007` <dbl>, `2008` <dbl>,
## #   `2009` <dbl>, `2010` <dbl>, `2011` <dbl>, `2012` <dbl>, `2013` <dbl>
```

6. Tidy and replicate the `us_rent_income` dataset in the `tidyr` package as given below.

```
## # A tibble: 14 x 5
##    GEOID NAME             moe income  rent
##    <chr> <chr>          <dbl>  <dbl> <dbl>
##  1 01    Alabama            3     NA   747
##  2 39    Ohio               2     NA   764
##  3 40    Oklahoma           3     NA   766
##  4 18    Indiana            3     NA   782
##  5 55    Wisconsin          3     NA   813
##  6 26    Michigan           3     NA   824
##  7 37    North Carolina     3     NA   844
##  8 42    Pennsylvania       3     NA   885
##  9 13    Georgia            3     NA   927
## 10 17    Illinois           3     NA   952
## 11 48    Texas              2     NA   952
## 12 12    Florida            3     NA  1077
## 13 36    New York           3     NA  1194
## 14 06    California         3     NA  1358
```

7. Tidy and replicate the `relig_income` dataset in the `tidyr` package as given below.

```
## # A tibble: 18 x 3
##    religion                Income Count
##    <chr>                   <chr>  <dbl>
##  1 Hindu                   <$10k      1
##  2 Other World Religions   <$10k      5
##  3 Muslim                  <$10k      6
##  4 Other Christian         <$10k      9
##  5 Atheist                 <$10k     12
##  6 Orthodox                <$10k     13
##  7 Don't know/refused      <$10k     15
##  8 Jewish                  <$10k     19
##  9 Jehovah's Witness       <$10k     20
## 10 Other Faiths            <$10k     20
## 11 Agnostic                <$10k     27
## 12 Buddhist                <$10k     27
## 13 Mormon                  <$10k     29
## 14 Unaffiliated            <$10k    217
## 15 Historically Black Prot <$10k    228
## 16 Mainline Prot           <$10k    289
## 17 Catholic                <$10k    418
## 18 Evangelical Prot        <$10k    575
```

8. Tidy and replicate the `billboard` dataset in the `tidyr` package as given below.

```
## # A tibble: 4 x 81
##   artist track  year month   day   wk1   wk2   wk3   wk4   wk5   wk6   wk7   wk8
##   <chr>  <chr> <int> <int> <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Backs~ Show~  2000     1     1    74    62    55    25    16    14    12    10
## 2 Brock~ A Co~  2000     1     1    93    75    92    NA    NA    NA    NA    NA
## 3 Diffi~ The ~  2000     1     1    98   100   100    90    93    94    NA    NA
## 4 Joe    I Wa~  2000     1     1    94    86    69    50    41    33    32    28
## # i 68 more variables: wk9 <dbl>, wk10 <dbl>, wk11 <dbl>, wk12 <dbl>,
## #   wk13 <dbl>, wk14 <dbl>, wk15 <dbl>, wk16 <dbl>, wk17 <dbl>, wk18 <dbl>,
## #   wk19 <dbl>, wk20 <dbl>, wk21 <dbl>, wk22 <dbl>, wk23 <dbl>, wk24 <dbl>,
## #   wk25 <dbl>, wk26 <dbl>, wk27 <dbl>, wk28 <dbl>, wk29 <dbl>, wk30 <dbl>,
## #   wk31 <dbl>, wk32 <dbl>, wk33 <dbl>, wk34 <dbl>, wk35 <dbl>, wk36 <dbl>,
## #   wk37 <dbl>, wk38 <dbl>, wk39 <dbl>, wk40 <dbl>, wk41 <dbl>, wk42 <dbl>,
## #   wk43 <dbl>, wk44 <dbl>, wk45 <dbl>, wk46 <dbl>, wk47 <dbl>, wk48 <dbl>, ...
```

9. Tidy and replicate the `airlines` dataset in the `nycflights13` package as given below.

```
## # A tibble: 16 x 2
##    carrier airline
##    <chr>   <chr>
##  1 9E      Endeavor
##  2 AA      American
##  3 AS      Alaska
##  4 B6      JetBlue
##  5 DL      Delta
##  6 EV      ExpressJet
##  7 F9      Frontier
##  8 FL      AirTran
##  9 HA      Hawaiian
## 10 MQ      Envoy
## 11 OO      SkyWest
## 12 UA      United
## 13 US      US
## 14 VX      Virgin
## 15 WN      Southwest
## 16 YV      Mesa
```

base R

1. Run the following codes and explain why the value of `address1` is shown as `NA` while the `class(address1)` is `numeric`?

```
x1 <- "Presidency"
x2 <- "University"
x3 <- "Kolkata"
address <- c(x1, x2, x3)
address1 <- as.numeric(address)
address1
class(address1)
```

2. Explain the following codes and their outputs.

```r
a1 <- 12; class(a1); length(a1)
names(a1) <- 'Number'; names(a1)
```

3. Explain the following codes and their outputs.

```r
a2 <- matrix(1:9, nrow = 3)
colnames(a2) <- c("A", "B", "C")

a2[c(TRUE, FALSE, TRUE), c("B", "A")]
```

4. Explain the following codes and their outputs.

```r
month_levels <- c(
  "Jan", "Feb", "Mar", "Apr", "May", "Jun",
  "Jul", "Aug", "Sep", "Oct", "Nov", "Dec"
)
a4 <- factor(c("Dec", "Apr", "Jan", "Mar"), levels = month_levels)
a4
```

5. Explain the following codes and their outputs.

```r
a5 <- factor(c("high", "low", "medium", "medium", "high"), levels = c("low", "medium", "high"), ordered
a5
```

6. Explain the following codes and their outputs.

```r
library(gapminder)
filter(gapminder, continent %in% c("Asia", "Africa"))
```

7. In the `geom_bar()` function explain the difference between the use of `stat = "identity"` and `stat = "count"`.

8. Explain the following codes and their outputs.

```r
s <- 1:5
rating <- factor(s)

(rating <- factor(s, ordered = TRUE,
                  levels = s))
```

9. Write a code to print the following output.

```
## # A tibble: 2 x 2
##   `@gmail.com`  `:)`
##   <chr>        <dbl>
## 1 presi            0
## 2 econ             1
```

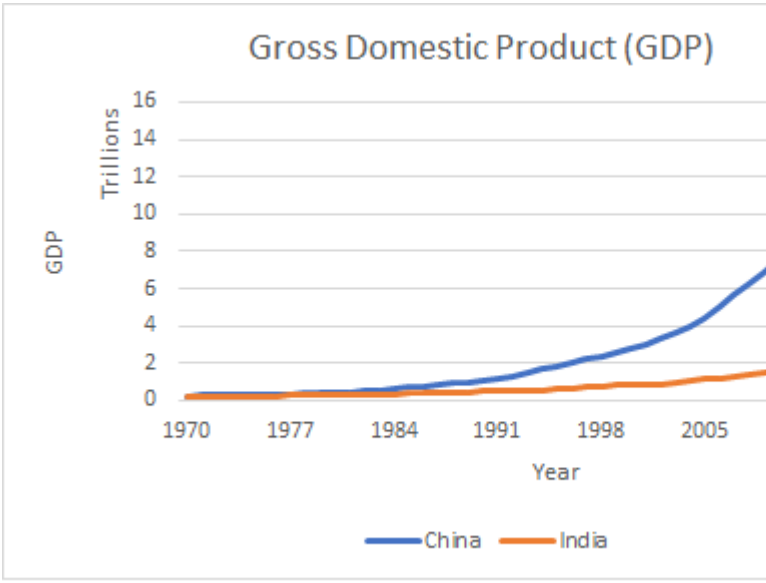
https://bookdown.org/sunboklee/ewha_r_2021_1/base-r-quiz.html#quiz-problem-1

Excel

1. Which specie has the longest and the widest petal in the `iris.xlsx` data?
2. What is the average displacement of a manual car with 4 cylinders in the `mtcars.xlsx` dataset?
3. Calculate the mean *mpg* (miles per gallon) of the cars with 6 and 4 cylinders in `mtcars.xlsx` dataset?
4. How many years of data are available for each country in the `GDP.xlsx` dataset? How many countries do not have data for all the years?
5. In the file `GDP.xlsx` how many countries do not have data on GDP?
6. Rank (without ties) the countries according to the Gross Domestic Product (GDP) in the `GDP.xlsx` dataset.
7. How does the number of cylinders `cyl` affect the mileage `mpg` for a given horsepower `hp`? Calculate the partial correlation in the `mtcars.xlsx` dataset.
8. How many missing values are there in total in the `GDP.xlsx` dataset? Find how many countries are listed in the data without using the filter option?

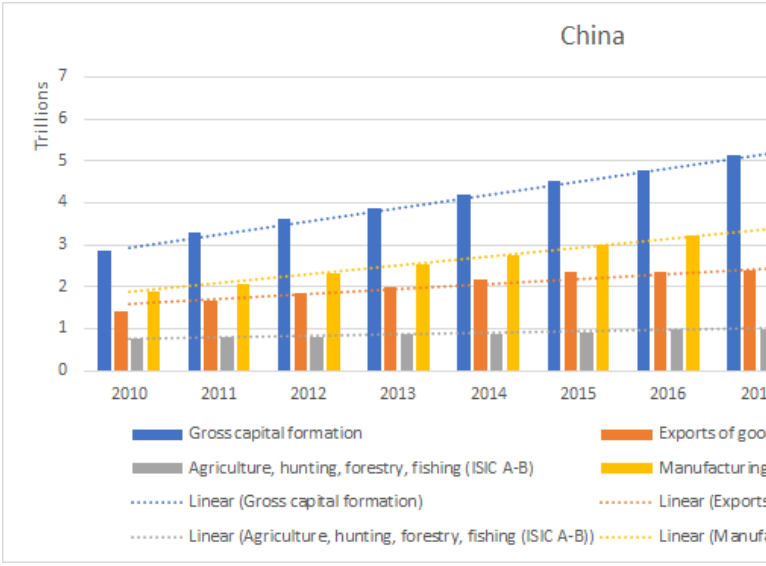
plots

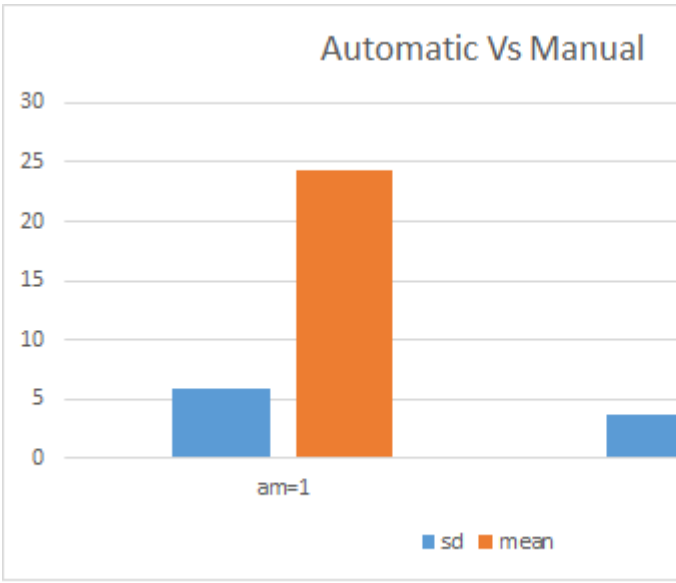
9. Using the data `GDP.xlsx`, for any two countries plot separate line charts for the components of GDP (Household Consumption Expenditure, Government final consumption expenditure, Net exports and Gross Capital formation).



10. Using the `GDP.xlsx` data replicate the following plot.

11. Using the `GDP.xlsx` data replicate the following plot.





12. Using the `mtcars.xlsx` dataset replicate the followuing plot.
13. Suppose that the firm's production function is $Q = F(K, L) = 50K^{0.5}L^{0.5}$. Suppose, too, that the price of labour w=5 and the price of capital r=20. What is the cost minimising input bundle if the firm wants to produce 1,000 units per year?