MarkdownNotes

FedericaFusi

March 09, 2021

Set up the workplace

Upload needed packages

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.4     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(tidytuesdayR)
```

Open your data

```
tuesdata <- tidytuesdayR::tt_load(2021, week = 2)
```

```
## --- Compiling #TidyTuesday Information for 2021-01-05 ----
```

```
## --- There is 1 file available ---
```

```
## --- Starting Download ---
```

```
##
##  Downloading file 1 of 1: `transit_cost.csv`
```

```
## --- Download complete ---
```

```
transit_cost <- tuesdata$transit_cost
```

Have a first look at the data

```
transit_cost
```

```
## # A tibble: 544 x 20
##        e country city  line  start_year end_year    rr length tunnel_per tunnel
##    <dbl> <chr>   <chr> <chr> <chr>      <chr>    <dbl>  <dbl> <chr>       <dbl>
##  1  7136 CA      Vanc~ Broa~ 2020       2025         0    5.7 87.72%        5
##  2  7137 CA      Toro~ Vaug~ 2009       2017         0    8.6 100.00%       8.6
##  3  7138 CA      Toro~ Scar~ 2020       2030         0    7.8 100.00%       7.8
##  4  7139 CA      Toro~ Onta~ 2020       2030         0   15.5 57.00%        8.8
##  5  7144 CA      Toro~ Yong~ 2020       2030         0    7.4 100.00%       7.4
##  6  7145 NL      Amst~ Nort~ 2003       2018         0    9.7 73.00%        7.1
```

```
##  7  7146 CA      Mont~ Blue~ 2020       2026         0    5.8 100.00%       5.8
##  8  7147 US      Seat~ U-Li~ 2009       2016         0    5.1 100.00%       5.1
##  9  7152 US      Los ~ Purp~ 2020       2027         0    4.2 100.00%       4.2
## 10  7153 US      Los ~ Purp~ 2018       2026         0    4.2 100.00%       4.2
## # ... with 534 more rows, and 10 more variables: stations <dbl>, source1 <chr>,
## #   cost <dbl>, currency <chr>, year <dbl>, ppp_rate <dbl>, real_cost <chr>,
## #   cost_km_millions <dbl>, source2 <chr>, reference <chr>
```

Check dimensions

```
dim(transit_cost) # 544 rows and 20 columns
```

```
## [1] 544  20
```

Cleaning

Variables of interest

I want to look at the relationship between **length of the project** and **real cost**. My scope is to *identify those projects whose cost per km is particularly high or low.*

```
class(transit_cost$real_cost) #character
```

```
## [1] "character"
```

```
class(transit_cost$length) #numeric
```

```
## [1] "numeric"
```

The variable **real_cost** needs to be converted into numeric. I quickly look at it to see potential issues that might arise with the conversation.

```
table(transit_cost$real_cost) #There seems to be a few "weird" entry that are character rather than num
```

```
##
##          0    1005.94 10122.0845    1015.92    1021.68   10234.85   10270.08
##          2          1          1          1          1          1          1
##     103.75     1032.3     1034.8       1044       1050        106    1065.56
##          1          1          1          1          1          1          1
##      10695    1072.60   10720.74    1080.58      10800    1084.21       1095
##          1          1          1          1          1          1          1
##       1100      11000   11006.45    1102.85    1105.66    11088.4       1115
##          1          1          1          1          1          1          1
##       1131    1141.82   11430.54     116.46    1164.84       1170    1170.65
##          1          1          1          1          1          1          1
## 11725.8655    1181.95       1183     1189.5    1194.67   11941.02       1200
##          1          1          1          1          1          1          1
##   12038.32     1207.7    1220.50    1227.46    1236.64      12400       1248
##          1          1          1          1          1          1          1
##    1255.49    1259.96    1265.85   12676.24   12709.89    1271.27     1283.4
##          1          1          1          1          1          1          1
##      12960  1298.6512   1307.625   13196.16    1333.15   13509.83       1352
##          1          1          1          1          1          1          1
##    1357.26     1380.4   13888.44     1409.2       1417     142.31   14396.76
##          1          1          1          1          1          1          1
##    1442.35   14426.65    1444.32     1445.6     1453.2     146.08    1462.55
##          1          1          1          1          1          1          1
##      14700     1489.8    149.736      15040       1514       1517    1519.04
```

```
##         1         1         1         1         1         1         1
##    1519.7   1520.47  15228.68   1531.93   1534.49    1541.8  15744.00
##         1         1         1         1         1         1         1
##   1580.22   1592.19      1600   1607.79   1618.08    1619.1      1620
##         1         1         1         1         1         1         1
##   1627.14   1635.33 1641.31476   1642.02      1656      1660   1669.60
##         1         1         1         1         1         1         1
##      1670  16819.52   1683.36   1705.04     17220    1726.4      1743
##         1         1         1         1         1         1         1
##      1756      1764   1775.13   1777.83    178.75   1808.08    1809.6
##         2         1         1         1         1         1         1
##    1828.6   1834.08      1850  1860.712   18659.2   1879.06    188.75
##         1         1         1         1         1         1         1
##      1890   1894.49  1905.648  19060.92  1920.148  1949.504   1954.37
##         1         1         1         1         1         1         1
##   19632.8    1968.5     197.5   1973.75   1979.19   1980.71   1981.35
##         1         1         1         1         1         1         1
##    199.57    1991.8     20040   2025.57    2036.7   20539.5   2109.81
##         1         1         1         1         1         1         1
##   2136.18     21371   2166.19    2169.6   218.875    218.89    2185.5
##         1         1         1         1         1         1         1
##      2187    2189.6   2205.71      2265   2269.12   2274.24      2289
##         1         1         1         1         1         1         1
##     231.4      2340    236.14    2377.2    240.89      2400   2403.11
##         1         1         1         1         1         3         1
##   2407.49   2422.58   2430.90    2432.5   2463.93   2465.46     24780
##         1         1         1         1         1         1         1
##   2479.98     248.3    251.55   2510.98   2568.99   2575.60      2592
##         1         1         1         1         1         1         1
##     25968   2619.01   2647.38   2653.07   2662.18      2688   2694.12
##         1         1         1         1         1         1         1
##   2747.02    2762.6     277.2   2770.47    2783.5     280.8      2800
##         1         1         1         1         1         1         1
##      2805     28250      2860     288.6 2934.30291   2962.36   2970.76
##         1         1         1         1         1         1         1
##      2976   2979.84   2981.31   2983.18    2985.4  2991.663      3000
##         1         1         1         1         1         1         2
##   3001.43   3004.04   3006.66   3010.96   3037.16     30400   3062.92
##         1         1         1         1         1         1         1
##   3063.46    3076.8    308.96   3126.15   3144.36      3150   3160.64
##         1         1         1         1         1         1         1
##   3168.21   3219.39   32514.6    3261.5   3283.77   3286.29    3286.3
##         1         1         1         1         1         1         1
##     333.5   3333.28      3339    335.62    3389.2    3400.8    3411.4
##         1         1         1         1         1         1         1
##   3428.17   3467.36   3469.86      3471   3491.19    3491.4   3495.66
##         1         1         1         1         1         1         1
##      3503   3523.74   3528.59   3557.74   3579.84       360      3600
##         1         1         1         1         1         1         1
##     362.7   3621.89   3632.44   3645.05   3658.90   3667.94   3676.05
##         1         1         1         1         1         1         1
##   3691.15   3699.57   3706.71     373.5    377.28   3776.04      3780
##         1         1         1         1         1         1         1
##     379.5      3795    3811.5      3819 38200.8626     383.5   3843.07
```

```
##         1         1         1         1         1         1         1
##   3859.08   3880.26   3905.71  3919.266   3925.04 3939.9768   3946.25
##         1         1         1         1         1         1         1
##    395.59   3955.31  3955.755       400     400.4   4008.91   4021.05
##         1         1         1         1         1         1         1
##      4030    4040.4    408.94   4088.96     409.5    411.62    4131.4
##         1         1         1         1         1         1         1
##   4133.72       416   4168.41   4176.74    419.22   4205.76   4206.77
##         1         2         1         1         1         1         1
##      4224    424.19   4251.33   4253.81   4256.45   4260.97    4281.3
##         1         1         1         1         1         1         1
##   4297.81   4333.65   4342.97   4358.99   4396.02   4398.48      4399
##         1         1         1         1         1         1         1
##   4414.84   4434.15  4436.208      4450   4462.63       451      4536
##         1         1         1         1         1         1         1
##   4552.64     45604   4585.15  4594.725      4620   4632.41      4646
##         1         1         1         1         1         1         1
##  4646.008    465.07   4650.10    4687.2      4704    4716.8    474.89
##         1         1         1         1         1         1         1
##    475.31   4762.59   4767.51  4767.822    478.83   4818.85   4818.87
##         1         1         1         1         1         1         1
##   4824.29   4836.82    485.55   4882.89      4900   4903.35      4930
##         1         1         1         1         1         1         1
##       494   4949.94   4982.69   4996.90   5040.02      5100     512.2
##         1         1         1         1         1         1         1
##   5133.84   5163.42     518.7   5180.32   5195.96     521.3   5214.56
##         1         1         1         1         1         1         1
##   5225.38    5241.6   5338.12   5344.17      5360   5381.01      5382
##         1         2         1         1         1         1         1
##   5383.68     540.5   5437.07  5455.264   5459.07    5480.5    5493.6
##         1         1         1         1         1         1         1
##   5509.77      5510    556.77       559   5627.97   5632.21   5647.00
##         1         1         1         1         1         1         1
##   5707.48   5729.79      5760  5806.647   5818.75       587   5885.37
##         1         1         2         1         1         1         1
##       595   5955.99      5969   5994.06      6000    603.45      6039
##         1         1         1         1         1         1         1
##      6048    6060.6   6098.42    616.32   6160.63    617.75   6174.12
##         1         1         1         1         1         1         1
##    618.72   6181.88   6214.01   6227.78       624   6276.80   6285.92
##         1         1         1         1         2         1         1
##   6327.04   6360.11   6369.81   6370.99      6390   6448.98   6461.56
##         1         1         1         1         1         1         1
##   6510.47 6524.7296    653.66   6579.83     66.25   661.248   6626.97
##         1         1         1         1         1         1         1
##  6631.368   6635.21     664.3   6640.14      6650    668.39    670.05
##         1         1         1         1         1         1         1
##       676   6771.08    680.44     682.5   6838.03      6840   6842.68
##         1         1         1         1         1         1         1
##    685.49     687.5    687.83   6879.16     69.72       690      6900
##         1         1         1         1         1         1         1
##    692.50   6930.45   6935.62   6950.14    696.32   6969.94   6980.47
##         1         1         1         1         1         1         1
##      70.2   7011.84   7031.77      7040       711   7159.91   7201.32
```

```
##          1          1          1          1          2          1          1
##      724.1    725.153    7310.38    7335.85      737.1    7389.11    7397.24
##          1          1          1          1          1          1          1
##    7401.19    7413.85    7459.06       7500    7539.44    7559.64       7560
##          1          1          1          1          1          1          1
##    7561.59       7590    7665.08     7740.2    7743.48        780    7903.42
##          1          1          1          1          1          1          1
##    7953.22      799.4    8023.07     808.75        810    8183.93    8369.93
##          1          1          1          1          1          1          1
##    8393.88     841.12     850.38     851.21    8665.39     867.09       87.5
##          1          1          1          1          1          1          1
##     870.58    8771.32    8790.68    8825.17     883.49     885.18    8861.14
##          1          1          1          1          1          1          1
##      892.5      895.7      90000     902.29     902.71     905.41    9103.64
##          1          1          1          1          1          1          1
##     911.52    9115.68    9120.06    9141.14     919.08     9219.2    9361.08
##          1          1          1          1          1          1          1
##       9400    9465.78     948.36       9500      955.5     962.36      964.6
##          1          1          1          1          1          1          1
##       9700        988        AVG     MEDIAN        MIN          N QUARTILE 1
##          1          1          1          1          1          1          1
## QUARTILE 3        STD
##          1          1
```

I am going to check out these unusual entries by looking at the data

```
View(transit_cost)
```

They seem to be some summary columns left in the dataset by mistake. I will remove them.

```
transit_cost2 =
  transit_cost %>%
  filter(transit_cost$real_cost != "AVG",
         transit_cost$real_cost != "MEDIAN",
         transit_cost$real_cost != "MIN",
         transit_cost$real_cost != "N",
         transit_cost$real_cost != "QUARTILE 1",
         transit_cost$real_cost != "QUARTILE 3",
         transit_cost$real_cost != "STD")

nrow(transit_cost) - nrow(transit_cost2)
```

```
## [1] 7
```

The pipe has correctly eliminated *7 observations*. The new dataset size is `nrow(transit_cost2)`.

I will check again the column to see if there are any other non-numeric data but everything seems fine.

```
table(transit_cost2$real_cost)
```

```
##
##          0    1005.94 10122.0845    1015.92    1021.68   10234.85   10270.08
##          2          1          1          1          1          1          1
##     103.75     1032.3     1034.8       1044       1050        106    1065.56
##          1          1          1          1          1          1          1
##      10695    1072.60   10720.74    1080.58      10800    1084.21       1095
##          1          1          1          1          1          1          1
##       1100      11000   11006.45    1102.85    1105.66    11088.4       1115
```

```
##          1          1          1          1          1          1          1
##       1131    1141.82   11430.54     116.46    1164.84       1170    1170.65
##          1          1          1          1          1          1          1
## 11725.8655    1181.95       1183     1189.5    1194.67   11941.02       1200
##          1          1          1          1          1          1          1
##   12038.32     1207.7    1220.50    1227.46    1236.64      12400       1248
##          1          1          1          1          1          1          1
##    1255.49    1259.96    1265.85   12676.24   12709.89    1271.27     1283.4
##          1          1          1          1          1          1          1
##      12960  1298.6512   1307.625   13196.16    1333.15   13509.83       1352
##          1          1          1          1          1          1          1
##    1357.26     1380.4   13888.44     1409.2       1417     142.31   14396.76
##          1          1          1          1          1          1          1
##    1442.35   14426.65    1444.32     1445.6     1453.2     146.08    1462.55
##          1          1          1          1          1          1          1
##      14700     1489.8    149.736      15040       1514       1517    1519.04
##          1          1          1          1          1          1          1
##     1519.7    1520.47   15228.68    1531.93    1534.49     1541.8   15744.00
##          1          1          1          1          1          1          1
##    1580.22    1592.19       1600    1607.79    1618.08     1619.1       1620
##          1          1          1          1          1          1          1
##    1627.14    1635.33 1641.31476    1642.02       1656       1660    1669.60
##          1          1          1          1          1          1          1
##       1670   16819.52    1683.36    1705.04      17220     1726.4       1743
##          1          1          1          1          1          1          1
##       1756       1764    1775.13    1777.83     178.75    1808.08     1809.6
##          2          1          1          1          1          1          1
##     1828.6    1834.08       1850   1860.712    18659.2    1879.06     188.75
##          1          1          1          1          1          1          1
##       1890    1894.49   1905.648   19060.92   1920.148   1949.504    1954.37
##          1          1          1          1          1          1          1
##    19632.8     1968.5      197.5    1973.75    1979.19    1980.71    1981.35
##          1          1          1          1          1          1          1
##     199.57     1991.8      20040    2025.57     2036.7    20539.5    2109.81
##          1          1          1          1          1          1          1
##    2136.18      21371    2166.19     2169.6    218.875     218.89     2185.5
##          1          1          1          1          1          1          1
##       2187     2189.6    2205.71       2265    2269.12    2274.24       2289
##          1          1          1          1          1          1          1
##      231.4       2340     236.14     2377.2     240.89       2400    2403.11
##          1          1          1          1          1          3          1
##    2407.49    2422.58    2430.90     2432.5    2463.93    2465.46      24780
##          1          1          1          1          1          1          1
##    2479.98      248.3     251.55    2510.98    2568.99    2575.60       2592
##          1          1          1          1          1          1          1
##      25968    2619.01    2647.38    2653.07    2662.18       2688    2694.12
##          1          1          1          1          1          1          1
##    2747.02     2762.6      277.2    2770.47     2783.5      280.8       2800
##          1          1          1          1          1          1          1
##       2805      28250       2860      288.6 2934.30291    2962.36    2970.76
##          1          1          1          1          1          1          1
##       2976    2979.84    2981.31    2983.18     2985.4   2991.663       3000
##          1          1          1          1          1          1          2
##    3001.43    3004.04    3006.66    3010.96    3037.16      30400    3062.92
```

```
##         1         1         1         1         1         1         1
##   3063.46    3076.8    308.96   3126.15   3144.36      3150   3160.64
##         1         1         1         1         1         1         1
##   3168.21   3219.39   32514.6    3261.5   3283.77   3286.29    3286.3
##         1         1         1         1         1         1         1
##     333.5   3333.28      3339    335.62    3389.2    3400.8    3411.4
##         1         1         1         1         1         1         1
##   3428.17   3467.36   3469.86      3471   3491.19    3491.4   3495.66
##         1         1         1         1         1         1         1
##      3503   3523.74   3528.59   3557.74   3579.84       360      3600
##         1         1         1         1         1         1         1
##     362.7   3621.89   3632.44   3645.05   3658.90   3667.94   3676.05
##         1         1         1         1         1         1         1
##   3691.15   3699.57   3706.71     373.5    377.28   3776.04      3780
##         1         1         1         1         1         1         1
##     379.5      3795    3811.5      3819 38200.8626     383.5   3843.07
##         1         1         1         1         1         1         1
##   3859.08   3880.26   3905.71  3919.266   3925.04 3939.9768   3946.25
##         1         1         1         1         1         1         1
##    395.59   3955.31  3955.755       400     400.4   4008.91   4021.05
##         1         1         1         1         1         1         1
##      4030    4040.4    408.94   4088.96     409.5    411.62    4131.4
##         1         1         1         1         1         1         1
##   4133.72       416   4168.41   4176.74    419.22   4205.76   4206.77
##         1         2         1         1         1         1         1
##      4224    424.19   4251.33   4253.81   4256.45   4260.97    4281.3
##         1         1         1         1         1         1         1
##   4297.81   4333.65   4342.97   4358.99   4396.02   4398.48      4399
##         1         1         1         1         1         1         1
##   4414.84   4434.15  4436.208      4450   4462.63       451      4536
##         1         1         1         1         1         1         1
##   4552.64     45604   4585.15  4594.725      4620   4632.41      4646
##         1         1         1         1         1         1         1
##  4646.008    465.07   4650.10    4687.2      4704    4716.8    474.89
##         1         1         1         1         1         1         1
##    475.31   4762.59   4767.51  4767.822    478.83   4818.85   4818.87
##         1         1         1         1         1         1         1
##   4824.29   4836.82    485.55   4882.89      4900   4903.35      4930
##         1         1         1         1         1         1         1
##       494   4949.94   4982.69   4996.90   5040.02      5100     512.2
##         1         1         1         1         1         1         1
##   5133.84   5163.42     518.7   5180.32   5195.96     521.3   5214.56
##         1         1         1         1         1         1         1
##   5225.38    5241.6   5338.12   5344.17      5360   5381.01      5382
##         1         2         1         1         1         1         1
##   5383.68     540.5   5437.07  5455.264   5459.07    5480.5    5493.6
##         1         1         1         1         1         1         1
##   5509.77      5510    556.77       559   5627.97   5632.21   5647.00
##         1         1         1         1         1         1         1
##   5707.48   5729.79      5760  5806.647   5818.75       587   5885.37
##         1         1         2         1         1         1         1
##       595   5955.99      5969   5994.06      6000    603.45      6039
##         1         1         1         1         1         1         1
##      6048    6060.6   6098.42    616.32   6160.63    617.75   6174.12
```

```
##          1          1          1          1          1          1          1
##     618.72    6181.88    6214.01    6227.78        624    6276.80    6285.92
##          1          1          1          1          2          1          1
##    6327.04    6360.11    6369.81    6370.99       6390    6448.98    6461.56
##          1          1          1          1          1          1          1
##    6510.47  6524.7296     653.66    6579.83      66.25    661.248    6626.97
##          1          1          1          1          1          1          1
##   6631.368    6635.21      664.3    6640.14       6650     668.39     670.05
##          1          1          1          1          1          1          1
##        676    6771.08     680.44      682.5    6838.03       6840    6842.68
##          1          1          1          1          1          1          1
##     685.49      687.5     687.83    6879.16      69.72        690       6900
##          1          1          1          1          1          1          1
##     692.50    6930.45    6935.62    6950.14     696.32    6969.94    6980.47
##          1          1          1          1          1          1          1
##       70.2    7011.84    7031.77       7040        711    7159.91    7201.32
##          1          1          1          1          2          1          1
##      724.1    725.153    7310.38    7335.85      737.1    7389.11    7397.24
##          1          1          1          1          1          1          1
##    7401.19    7413.85    7459.06       7500    7539.44    7559.64       7560
##          1          1          1          1          1          1          1
##    7561.59       7590    7665.08     7740.2    7743.48        780    7903.42
##          1          1          1          1          1          1          1
##    7953.22      799.4    8023.07     808.75        810    8183.93    8369.93
##          1          1          1          1          1          1          1
##    8393.88     841.12     850.38     851.21    8665.39     867.09       87.5
##          1          1          1          1          1          1          1
##     870.58    8771.32    8790.68    8825.17     883.49     885.18    8861.14
##          1          1          1          1          1          1          1
##      892.5      895.7      90000     902.29     902.71     905.41    9103.64
##          1          1          1          1          1          1          1
##     911.52    9115.68    9120.06    9141.14     919.08     9219.2    9361.08
##          1          1          1          1          1          1          1
##       9400    9465.78     948.36       9500      955.5     962.36      964.6
##          1          1          1          1          1          1          1
##       9700        988
##          1          1
```

I will convert the variable into numeric

```
transit_cost2$real_cost2 = as.numeric(transit_cost2$real_cost)
```

Check missing values in both variables

```
transit_cost2 %>%
  summarise(sum(is.na(real_cost2)),
            sum(is.na(length)))
```

```
## # A tibble: 1 x 2
##   `sum(is.na(real_cost2))` `sum(is.na(length))`
##                      <int>                <int>
## 1                        0                    0
```

There aren't any, so we are good to start the plot

A quick note

Note what happens when we convert the original column from the original dataset

```
transit_cost$real_cost2 = as.numeric(transit_cost$real_cost)
```

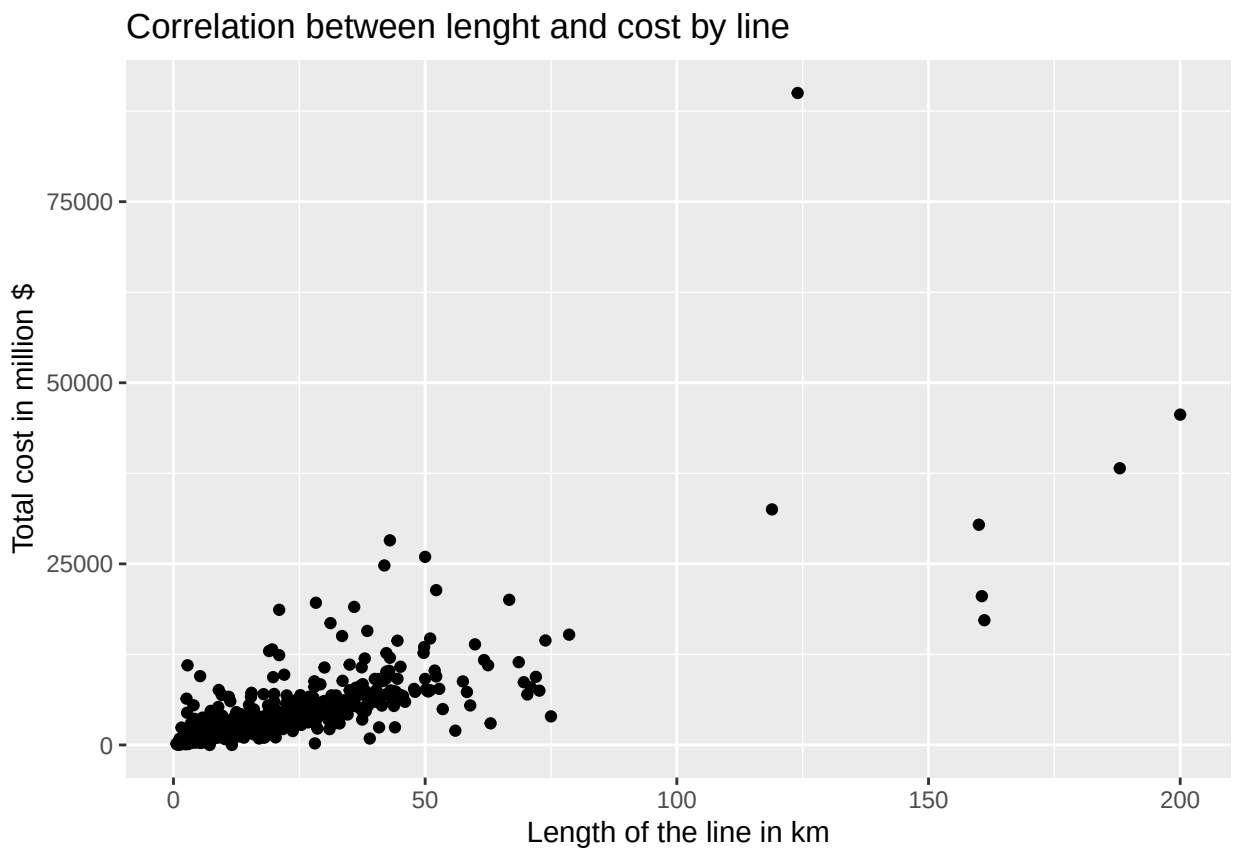
```
## Warning: NAs introduced by coercion
```

We get a warning message **Warning message: NAs introduced by coercion**. This a warning message that it is worth checking out. At a minimum you might want to inspect the column manually to see what is going on.

My scatterplot

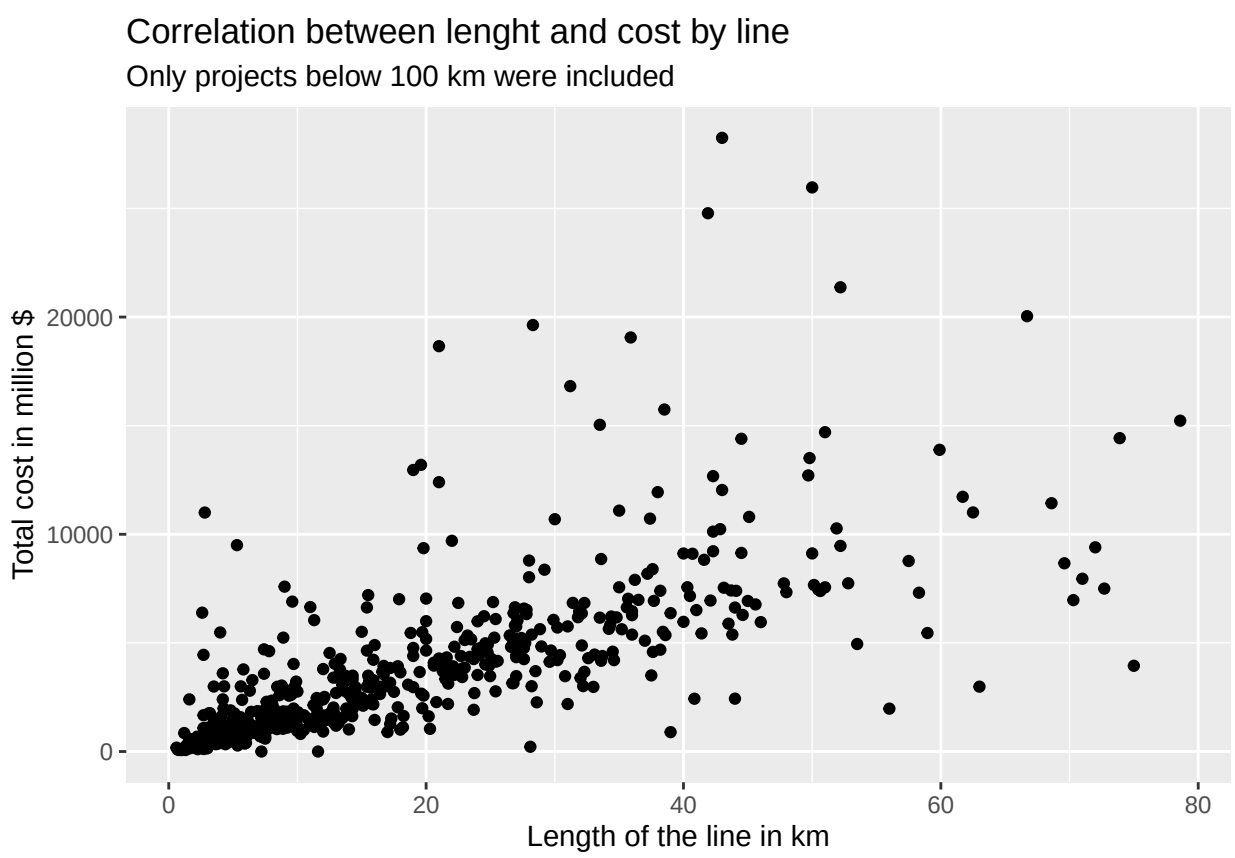
I will start with simple scatterplot to see how observations look like.

```
transit_cost2 %>%
  ggplot() +
  geom_point(mapping = aes(y = real_cost2,
                           x = length)) +
  xlab("Length of the line in km") +
  ylab("Total cost in million $") +
  labs(title = "Correlation between lenght and cost by line")
```



There are **seven outliers** whose length of the line is greater than 100 km. These outliers reduce our attention on the vast majority of the project so I decided to remove them. I need to remember to report this information in the graph.

```
transit_cost2 %>%
  filter(length <=100) %>% # Eliminate the outliers
  ggplot() +
  geom_point(mapping = aes(y = real_cost2,
                           x = length)) +
  xlab("Length of the line in km") +
  ylab("Total cost in million $") +
  labs(title = "Correlation between lenght and cost by line",
       subtitle = "Only projects below 100 km were included")
```



The graph looks much better now. I want to highlight those projects that are particularly expensive or particularly cheap. I decide to identify them as *being in the top 10 or bottom 10 quantile* considering the cost per km.

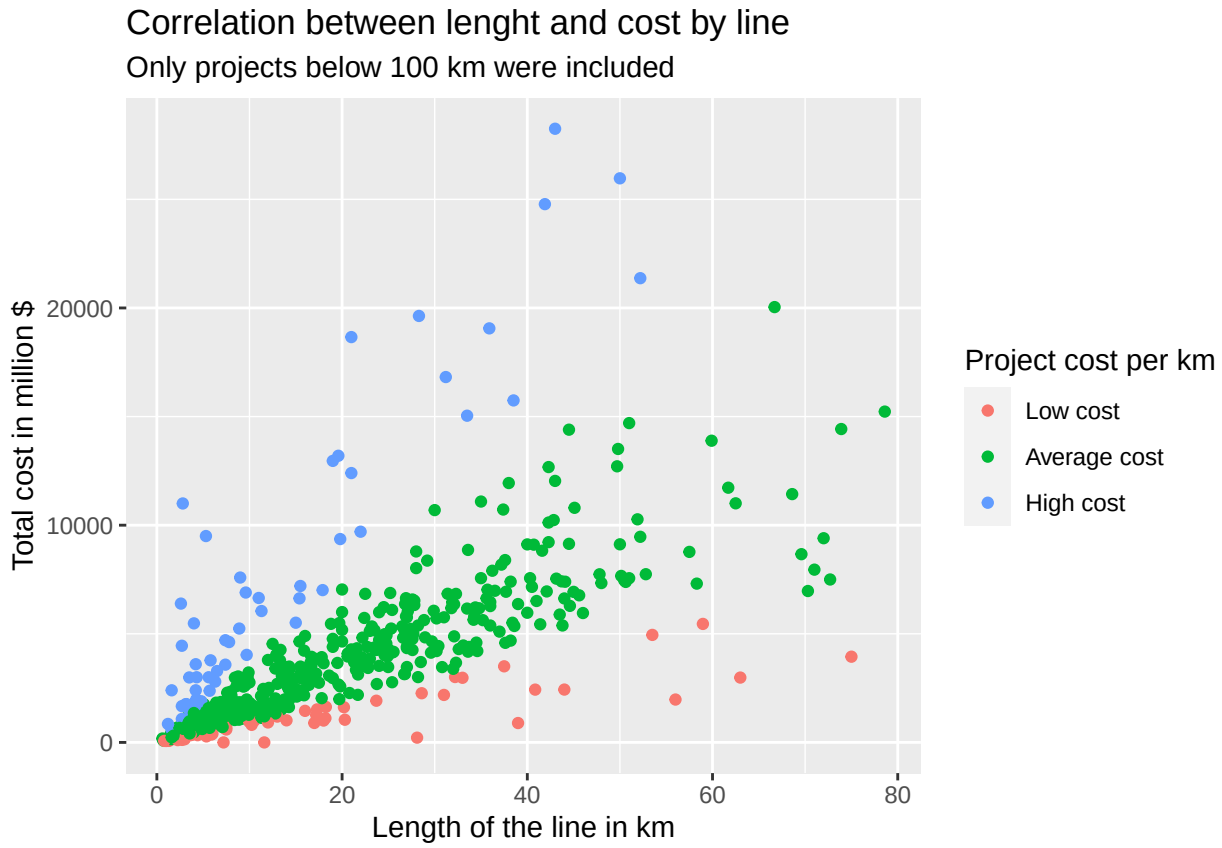
```
transit_cost2 %>%
  filter(length <= 100) %>%

# I need to create a new variable identify the ratio cost/length
mutate(cost_km = real_cost2 / length,

# I now create a new dummy variable indicating whether the ratio is in the top 10 (=2), bottom ten (=0)
# I categorize them 0,1,2 so that the label will be ordered as low, average, high cost
         ratio_cost = ifelse(cost_km > quantile(cost_km, 0.90), 2,
                             ifelse(cost_km < quantile(cost_km, 0.10), 0, 1))) %>%

# Now I start my plot. Same as before but I add color in the aes function to color my dots based on the
```

```
  ggplot() +
  geom_point(mapping = aes(y = real_cost2,
                           x = length,
                           color = as.character(ratio_cost))) + #Remember to change this as.character o
  xlab("Length of the line in km") +
  ylab("Total cost in million $") +
  labs(title = "Correlation between lenght and cost by line",
       subtitle = "Only projects below 100 km were included",
    # I change the legend label
       color = "Project cost per km") +
    # I change the lables of the different category using a scale. Note this is _discrete because I con
  scale_color_discrete(labels = c("Low cost", "Average cost", "High cost"))
```



Since I am pretty satisfied by the skeleton of my scatterplot, I can now change the graphic a bit. I want to change the color to highligh low and high cost projects and I want to add more brakes in my y_axis.

```
transit_cost2 %>%
  filter(length <= 100) %>%

  mutate(cost_km = real_cost2 / length,
         ratio_cost = ifelse(cost_km > quantile(cost_km, 0.90), 2,
                             ifelse(cost_km < quantile(cost_km, 0.10), 0, 1))) %>%
  ggplot() +
  geom_point(mapping = aes(y = real_cost2,
                           x = length,
                           fill = as.character(ratio_cost),
```

```
                          color = as.character(ratio_cost)),
             size = 2,
             shape = 21,
             alpha = 0.5) +
xlab("Length of the line in km") +
ylab("Total cost in million $") +
labs(title = "Correlation between lenght and cost by line",
     subtitle = "Only projects below 100 km were included",
     color = "Project cost per km") +
# Changing color to make
scale_fill_manual(values = c("#d62828", "#ced4da", "#003049"),
                  labels = c("Low cost", "Average cost", "High cost")) +
scale_color_manual(values = c("#d62828", "#ced4da", "#003049")) +
scale_y_continuous(breaks = seq(0, 30000, by = 5000))
```