title: “Data Transformation with dplyr Exercises” author: “Hadley Wickham, Garrett Grolemund, and Johnathan Clementi” date: “2/23/2020” output: word\_document —

These are exercises based on the exercises in R for Data Science by Wickham and Grolemund.

The website where the book appears is (and will always be) **free to use**, and is licensed under the [Creative Commons Attribution-NonCommercial-NoDerivs 3.0](http://creativecommons.org/licenses/by-nc-nd/3.0/us/) License.

# Data transformation

We will use the data set fueleconomy:vehicles

library(tidyverse)

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

## v ggplot2 3.2.1 v purrr 0.3.3  
## v tibble 2.1.3 v dplyr 0.8.3  
## v tidyr 1.0.2 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.4.0

## -- Conflicts -------------------------------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(fueleconomy)  
vehicles<-fueleconomy::vehicles

from <https://cran.r-project.org/web/packages/fueleconomy/fueleconomy.pdf>

vehicles- 5 cat., 4 quant, 1 case id id Unique EPA identifier make Manufacturer model Model name year Model year class EPA vehicle size class, <http://www.fueleconomy.gov/feg/ws/wsData.shtml#VClass> trans Transmission drive Drive train cyl Number of cylinders displ Engine displacement, in litres fuel Fuel type hwy Highway fuel economy, in mpg cty City fuel economy, in mpg

vehicles

## # A tibble: 33,442 x 12  
## id make model year class trans drive cyl displ fuel hwy cty  
## <int> <chr> <chr> <int> <chr> <chr> <chr> <int> <dbl> <chr> <int> <int>  
## 1 27550 AM Ge~ DJ Po ~ 1984 Specia~ Auto~ 2-Whe~ 4 2.5 Regu~ 17 18  
## 2 28426 AM Ge~ DJ Po ~ 1984 Specia~ Auto~ 2-Whe~ 4 2.5 Regu~ 17 18  
## 3 27549 AM Ge~ FJ8c P~ 1984 Specia~ Auto~ 2-Whe~ 6 4.2 Regu~ 13 13  
## 4 28425 AM Ge~ FJ8c P~ 1984 Specia~ Auto~ 2-Whe~ 6 4.2 Regu~ 13 13  
## 5 1032 AM Ge~ Post O~ 1985 Specia~ Auto~ Rear-~ 4 2.5 Regu~ 17 16  
## 6 1033 AM Ge~ Post O~ 1985 Specia~ Auto~ Rear-~ 6 4.2 Regu~ 13 13  
## 7 3347 ASC I~ GNX 1987 Midsiz~ Auto~ Rear-~ 6 3.8 Prem~ 21 14  
## 8 13309 Acura 2.2CL/~ 1997 Subcom~ Auto~ Front~ 4 2.2 Regu~ 26 20  
## 9 13310 Acura 2.2CL/~ 1997 Subcom~ Manu~ Front~ 4 2.2 Regu~ 28 22  
## 10 13311 Acura 2.2CL/~ 1997 Subcom~ Auto~ Front~ 6 3 Regu~ 26 18  
## # ... with 33,432 more rows

count(vehicles, make)

## # A tibble: 128 x 2  
## make n  
## <chr> <int>  
## 1 Acura 269  
## 2 Alfa Romeo 39  
## 3 AM General 6  
## 4 American Motors Corporation 27  
## 5 ASC Incorporated 1  
## 6 Aston Martin 112  
## 7 Audi 772  
## 8 Aurora Cars Ltd 1  
## 9 Autokraft Limited 4  
## 10 Azure Dynamics 2  
## # ... with 118 more rows

count(vehicles, model)

## # A tibble: 3,198 x 2  
## model n  
## <chr> <int>  
## 1 1-Ton Truck 2WD 16  
## 2 100 12  
## 3 100 quattro 9  
## 4 100 quattro Wagon 3  
## 5 100 Wagon 3  
## 6 1000 13  
## 7 128ci Convertible 8  
## 8 128i 12  
## 9 128i Convertible 4  
## 10 135i 12  
## # ... with 3,188 more rows

count(vehicles, year)

## # A tibble: 32 x 2  
## year n  
## <int> <int>  
## 1 1984 784  
## 2 1985 1701  
## 3 1986 1210  
## 4 1987 1247  
## 5 1988 1130  
## 6 1989 1153  
## 7 1990 1078  
## 8 1991 1132  
## 9 1992 1121  
## 10 1993 1093  
## # ... with 22 more rows

count(vehicles, class)

## # A tibble: 34 x 2  
## class n  
## <chr> <int>  
## 1 Compact Cars 4739  
## 2 Large Cars 1533  
## 3 Midsize-Large Station Wagons 627  
## 4 Midsize Cars 3621  
## 5 Midsize Station Wagons 415  
## 6 Minicompact Cars 1080  
## 7 Minivan - 2WD 308  
## 8 Minivan - 4WD 44  
## 9 Small Pickup Trucks 538  
## 10 Small Pickup Trucks 2WD 392  
## # ... with 24 more rows

count(vehicles, trans)

## # A tibble: 48 x 2  
## trans n  
## <chr> <int>  
## 1 "" 8  
## 2 "Auto (AV-S6)" 1  
## 3 "Auto (AV-S8)" 1  
## 4 "Auto (AV)" 2  
## 5 "Auto(A1)" 1  
## 6 "Auto(A8)" 1  
## 7 "Auto(AM-S6)" 43  
## 8 "Auto(AM-S7)" 93  
## 9 "Auto(AM-S8)" 1  
## 10 "Auto(AM5)" 10  
## # ... with 38 more rows

count(vehicles, drive)

## # A tibble: 7 x 2  
## drive n  
## <chr> <int>  
## 1 2-Wheel Drive 507  
## 2 4-Wheel Drive 699  
## 3 4-Wheel or All-Wheel Drive 6647  
## 4 All-Wheel Drive 1267  
## 5 Front-Wheel Drive 12233  
## 6 Part-time 4-Wheel Drive 96  
## 7 Rear-Wheel Drive 11993

count(vehicles, cyl)

## # A tibble: 10 x 2  
## cyl n  
## <int> <int>  
## 1 2 45  
## 2 3 182  
## 3 4 12381  
## 4 5 718  
## 5 6 11885  
## 6 8 7550  
## 7 10 138  
## 8 12 478  
## 9 16 7  
## 10 NA 58

count(vehicles, displ)

## # A tibble: 65 x 2  
## displ n  
## <dbl> <int>  
## 1 0 3  
## 2 1 157  
## 3 1.1 8  
## 4 1.2 35  
## 5 1.3 178  
## 6 1.4 80  
## 7 1.5 610  
## 8 1.6 1264  
## 9 1.7 50  
## 10 1.8 1353  
## # ... with 55 more rows

count(vehicles, fuel)

## # A tibble: 13 x 2  
## fuel n  
## <chr> <int>  
## 1 CNG 58  
## 2 Diesel 874  
## 3 Electricity 55  
## 4 Gasoline or E85 1043  
## 5 Gasoline or natural gas 18  
## 6 Gasoline or propane 8  
## 7 Midgrade 43  
## 8 Premium 8617  
## 9 Premium and Electricity 1  
## 10 Premium Gas or Electricity 7  
## 11 Premium or E85 88  
## 12 Regular 22622  
## 13 Regular Gas and Electricity 8

count(vehicles, hwy)

## # A tibble: 68 x 2  
## hwy n  
## <int> <int>  
## 1 9 13  
## 2 10 66  
## 3 11 62  
## 4 12 275  
## 5 13 295  
## 6 14 453  
## 7 15 847  
## 8 16 1257  
## 9 17 2094  
## 10 18 1547  
## # ... with 58 more rows

count(vehicles, cty)

## # A tibble: 67 x 2  
## cty n  
## <int> <int>  
## 1 6 5  
## 2 7 24  
## 3 8 88  
## 4 9 196  
## 5 10 527  
## 6 11 1403  
## 7 12 1484  
## 8 13 1873  
## 9 14 2685  
## 10 15 4020  
## # ... with 57 more rows

### Exercises

1. Find all vehicles that
   1. were made after 2010
   2. were made by Ford or General Motors
   3. had 2-Wheel Drive or Front-Wheel Driv
   4. had a manual transmition of any kind.
   5. had city miles per gallon above the highways miles per gallon.
   6. had highway miles per gallon above 25 and less than 30.
   7. had displacement between 1 and 5 (inclusive)

1A.

#a.  
 filter(vehicles, year > 2010)

## # A tibble: 4,886 x 12  
## id make model year class trans drive cyl displ fuel hwy cty  
## <int> <chr> <chr> <int> <chr> <chr> <chr> <int> <dbl> <chr> <int> <int>  
## 1 33146 Acura ILX 2013 Compact~ Autom~ Front~ 4 2 Prem~ 35 24  
## 2 33147 Acura ILX 2013 Compact~ Manua~ Front~ 4 2.4 Prem~ 31 22  
## 3 33566 Acura ILX 2014 Compact~ Autom~ Front~ 4 2 Prem~ 35 24  
## 4 33567 Acura ILX 2014 Compact~ Manua~ Front~ 4 2.4 Prem~ 31 22  
## 5 33148 Acura ILX H~ 2013 Compact~ Auto(~ Front~ 4 1.5 Prem~ 38 39  
## 6 33607 Acura ILX H~ 2014 Compact~ Auto(~ Front~ 4 1.5 Prem~ 38 39  
## 7 33590 Acura MDX 2~ 2014 Small S~ Autom~ Front~ 6 3.5 Prem~ 28 20  
## 8 30772 Acura MDX 4~ 2011 Sport U~ Autom~ All-W~ 6 3.7 Prem~ 21 16  
## 9 31947 Acura MDX 4~ 2012 Sport U~ Autom~ All-W~ 6 3.7 Prem~ 21 16  
## 10 33234 Acura MDX 4~ 2013 Small S~ Autom~ All-W~ 6 3.7 Prem~ 21 16  
## # ... with 4,876 more rows

#b.   
 filter(vehicles, make == 'Ford' | make == 'General Motors')

## # A tibble: 2,808 x 12  
## id make model year class trans drive cyl displ fuel hwy cty  
## <int> <chr> <chr> <int> <chr> <chr> <chr> <int> <dbl> <chr> <int> <int>  
## 1 2638 Ford Aerost~ 1986 Vans Automa~ Rear-W~ 4 2.3 Regu~ 22 18  
## 2 2639 Ford Aerost~ 1986 Vans Manual~ Rear-W~ 4 2.3 Regu~ 26 23  
## 3 2640 Ford Aerost~ 1986 Vans Automa~ Rear-W~ 6 2.8 Regu~ 21 15  
## 4 2641 Ford Aerost~ 1986 Vans Manual~ Rear-W~ 6 2.8 Regu~ 22 16  
## 5 2642 Ford Aerost~ 1986 Vans Manual~ Rear-W~ 6 3 Regu~ 22 17  
## 6 2643 Ford Aerost~ 1986 Vans Automa~ Rear-W~ 6 3 Regu~ 22 15  
## 7 3883 Ford Aerost~ 1987 Vans Automa~ Rear-W~ 4 2.3 Regu~ 24 18  
## 8 3884 Ford Aerost~ 1987 Vans Manual~ Rear-W~ 4 2.3 Regu~ 26 23  
## 9 3885 Ford Aerost~ 1987 Vans Automa~ Rear-W~ 6 3 Regu~ 22 16  
## 10 3886 Ford Aerost~ 1987 Vans Manual~ Rear-W~ 6 3 Regu~ 22 17  
## # ... with 2,798 more rows

#c.  
 filter(vehicles, drive == '2-Wheel Drive' | drive == 'Front-Wheel Drive')

## # A tibble: 12,740 x 12  
## id make model year class trans drive cyl displ fuel hwy cty  
## <int> <chr> <chr> <int> <chr> <chr> <chr> <int> <dbl> <chr> <int> <int>  
## 1 27550 AM Ge~ DJ Po ~ 1984 Specia~ Auto~ 2-Whe~ 4 2.5 Regu~ 17 18  
## 2 28426 AM Ge~ DJ Po ~ 1984 Specia~ Auto~ 2-Whe~ 4 2.5 Regu~ 17 18  
## 3 27549 AM Ge~ FJ8c P~ 1984 Specia~ Auto~ 2-Whe~ 6 4.2 Regu~ 13 13  
## 4 28425 AM Ge~ FJ8c P~ 1984 Specia~ Auto~ 2-Whe~ 6 4.2 Regu~ 13 13  
## 5 13309 Acura 2.2CL/~ 1997 Subcom~ Auto~ Front~ 4 2.2 Regu~ 26 20  
## 6 13310 Acura 2.2CL/~ 1997 Subcom~ Manu~ Front~ 4 2.2 Regu~ 28 22  
## 7 13311 Acura 2.2CL/~ 1997 Subcom~ Auto~ Front~ 6 3 Regu~ 26 18  
## 8 14038 Acura 2.3CL/~ 1998 Subcom~ Auto~ Front~ 4 2.3 Regu~ 27 19  
## 9 14039 Acura 2.3CL/~ 1998 Subcom~ Manu~ Front~ 4 2.3 Regu~ 29 21  
## 10 14040 Acura 2.3CL/~ 1998 Subcom~ Auto~ Front~ 6 3 Regu~ 26 17  
## # ... with 12,730 more rows

#d.  
 filter(vehicles, grepl('Manual', trans))

## # A tibble: 11,396 x 12  
## id make model year class trans drive cyl displ fuel hwy cty  
## <int> <chr> <chr> <int> <chr> <chr> <chr> <int> <dbl> <chr> <int> <int>  
## 1 13310 Acura 2.2CL~ 1997 Subcom~ Manual~ Front~ 4 2.2 Regu~ 28 22  
## 2 14039 Acura 2.3CL~ 1998 Subcom~ Manual~ Front~ 4 2.3 Regu~ 29 21  
## 3 14835 Acura 2.3CL~ 1999 Subcom~ Manual~ Front~ 4 2.3 Regu~ 29 21  
## 4 18458 Acura 3.2CL 2003 Compac~ Manual~ Front~ 6 3.2 Prem~ 26 17  
## 5 33147 Acura ILX 2013 Compac~ Manual~ Front~ 4 2.4 Prem~ 31 22  
## 6 33567 Acura ILX 2014 Compac~ Manual~ Front~ 4 2.4 Prem~ 31 22  
## 7 1834 Acura Integ~ 1986 Subcom~ Manual~ Front~ 4 1.6 Regu~ 28 23  
## 8 3038 Acura Integ~ 1987 Subcom~ Manual~ Front~ 4 1.6 Regu~ 28 23  
## 9 4184 Acura Integ~ 1988 Subcom~ Manual~ Front~ 4 1.6 Regu~ 28 23  
## 10 5304 Acura Integ~ 1989 Subcom~ Manual~ Front~ 4 1.6 Regu~ 28 23  
## # ... with 11,386 more rows

# Used grepl for regex insteal of typing out each name... which would look like:  
   
 filter(vehicles, trans == 'Manual 3-spd' | trans == 'Manual 4-spd' | trans == 'Manual 5-spd' | trans == 'Manual 6-spd' | trans == 'Manual 7-spd' | trans == 'Manual(M7)')

## # A tibble: 11,395 x 12  
## id make model year class trans drive cyl displ fuel hwy cty  
## <int> <chr> <chr> <int> <chr> <chr> <chr> <int> <dbl> <chr> <int> <int>  
## 1 13310 Acura 2.2CL~ 1997 Subcom~ Manual~ Front~ 4 2.2 Regu~ 28 22  
## 2 14039 Acura 2.3CL~ 1998 Subcom~ Manual~ Front~ 4 2.3 Regu~ 29 21  
## 3 14835 Acura 2.3CL~ 1999 Subcom~ Manual~ Front~ 4 2.3 Regu~ 29 21  
## 4 18458 Acura 3.2CL 2003 Compac~ Manual~ Front~ 6 3.2 Prem~ 26 17  
## 5 33147 Acura ILX 2013 Compac~ Manual~ Front~ 4 2.4 Prem~ 31 22  
## 6 33567 Acura ILX 2014 Compac~ Manual~ Front~ 4 2.4 Prem~ 31 22  
## 7 1834 Acura Integ~ 1986 Subcom~ Manual~ Front~ 4 1.6 Regu~ 28 23  
## 8 3038 Acura Integ~ 1987 Subcom~ Manual~ Front~ 4 1.6 Regu~ 28 23  
## 9 4184 Acura Integ~ 1988 Subcom~ Manual~ Front~ 4 1.6 Regu~ 28 23  
## 10 5304 Acura Integ~ 1989 Subcom~ Manual~ Front~ 4 1.6 Regu~ 28 23  
## # ... with 11,385 more rows

#e.  
 filter(vehicles, cty > hwy)

## # A tibble: 166 x 12  
## id make model year class trans drive cyl displ fuel hwy cty  
## <int> <chr> <chr> <int> <chr> <chr> <chr> <int> <dbl> <chr> <int> <int>  
## 1 27550 AM Ge~ DJ Po ~ 1984 Specia~ Autom~ 2-Wh~ 4 2.5 Regu~ 17 18  
## 2 28426 AM Ge~ DJ Po ~ 1984 Specia~ Autom~ 2-Wh~ 4 2.5 Regu~ 17 18  
## 3 33148 Acura ILX Hy~ 2013 Compac~ Auto(~ Fron~ 4 1.5 Prem~ 38 39  
## 4 33607 Acura ILX Hy~ 2014 Compac~ Auto(~ Fron~ 4 1.5 Prem~ 38 39  
## 5 32276 BMW Active~ 2011 Subcom~ Autom~ Rear~ NA NA Elec~ 96 107  
## 6 32277 CODA ~ CODA 2012 Subcom~ Autom~ Fron~ NA NA Elec~ 68 77  
## 7 33337 CODA ~ CODA 2013 Subcom~ Autom~ Fron~ NA NA Elec~ 68 77  
## 8 33640 Chevr~ Spark ~ 2014 Subcom~ Autom~ Fron~ NA NA Elec~ 109 128  
## 9 30973 Dodge Carava~ 1999 Miniva~ Autom~ 2-Wh~ NA NA Elec~ 33 37  
## 10 33396 Fiat 500e 2013 Minico~ Autom~ Fron~ NA NA Elec~ 108 122  
## # ... with 156 more rows

# Important: still contains NA values...  
  
 #f.  
 filter(vehicles, hwy > 25 & hwy < 30)

## # A tibble: 7,089 x 12  
## id make model year class trans drive cyl displ fuel hwy cty  
## <int> <chr> <chr> <int> <chr> <chr> <chr> <int> <dbl> <chr> <int> <int>  
## 1 13309 Acura 2.2CL~ 1997 Subcom~ Automa~ Front~ 4 2.2 Regu~ 26 20  
## 2 13310 Acura 2.2CL~ 1997 Subcom~ Manual~ Front~ 4 2.2 Regu~ 28 22  
## 3 13311 Acura 2.2CL~ 1997 Subcom~ Automa~ Front~ 6 3 Regu~ 26 18  
## 4 14038 Acura 2.3CL~ 1998 Subcom~ Automa~ Front~ 4 2.3 Regu~ 27 19  
## 5 14039 Acura 2.3CL~ 1998 Subcom~ Manual~ Front~ 4 2.3 Regu~ 29 21  
## 6 14040 Acura 2.3CL~ 1998 Subcom~ Automa~ Front~ 6 3 Regu~ 26 17  
## 7 14834 Acura 2.3CL~ 1999 Subcom~ Automa~ Front~ 4 2.3 Regu~ 27 20  
## 8 14835 Acura 2.3CL~ 1999 Subcom~ Manual~ Front~ 4 2.3 Regu~ 29 21  
## 9 14836 Acura 2.3CL~ 1999 Subcom~ Automa~ Front~ 6 3 Regu~ 26 17  
## 10 16573 Acura 3.2CL 2001 Compac~ Automa~ Front~ 6 3.2 Prem~ 27 17  
## # ... with 7,079 more rows

#g.  
 filter(vehicles, displ >= 1 & displ <= 5)

## # A tibble: 28,686 x 12  
## id make model year class trans drive cyl displ fuel hwy cty  
## <int> <chr> <chr> <int> <chr> <chr> <chr> <int> <dbl> <chr> <int> <int>  
## 1 27550 AM Ge~ DJ Po ~ 1984 Specia~ Auto~ 2-Whe~ 4 2.5 Regu~ 17 18  
## 2 28426 AM Ge~ DJ Po ~ 1984 Specia~ Auto~ 2-Whe~ 4 2.5 Regu~ 17 18  
## 3 27549 AM Ge~ FJ8c P~ 1984 Specia~ Auto~ 2-Whe~ 6 4.2 Regu~ 13 13  
## 4 28425 AM Ge~ FJ8c P~ 1984 Specia~ Auto~ 2-Whe~ 6 4.2 Regu~ 13 13  
## 5 1032 AM Ge~ Post O~ 1985 Specia~ Auto~ Rear-~ 4 2.5 Regu~ 17 16  
## 6 1033 AM Ge~ Post O~ 1985 Specia~ Auto~ Rear-~ 6 4.2 Regu~ 13 13  
## 7 3347 ASC I~ GNX 1987 Midsiz~ Auto~ Rear-~ 6 3.8 Prem~ 21 14  
## 8 13309 Acura 2.2CL/~ 1997 Subcom~ Auto~ Front~ 4 2.2 Regu~ 26 20  
## 9 13310 Acura 2.2CL/~ 1997 Subcom~ Manu~ Front~ 4 2.2 Regu~ 28 22  
## 10 13311 Acura 2.2CL/~ 1997 Subcom~ Auto~ Front~ 6 3 Regu~ 26 18  
## # ... with 28,676 more rows

1. Another useful dplyr filtering helper is between(). What does it do? Can you use it to simplify the code needed to answer the previous part?

2A. The between() function filters data vector that fall between the left and right bounds (inclusive). It is useful for simplifying code.

filter(vehicles, between(hwy, 26, 29))

## # A tibble: 7,089 x 12  
## id make model year class trans drive cyl displ fuel hwy cty  
## <int> <chr> <chr> <int> <chr> <chr> <chr> <int> <dbl> <chr> <int> <int>  
## 1 13309 Acura 2.2CL~ 1997 Subcom~ Automa~ Front~ 4 2.2 Regu~ 26 20  
## 2 13310 Acura 2.2CL~ 1997 Subcom~ Manual~ Front~ 4 2.2 Regu~ 28 22  
## 3 13311 Acura 2.2CL~ 1997 Subcom~ Automa~ Front~ 6 3 Regu~ 26 18  
## 4 14038 Acura 2.3CL~ 1998 Subcom~ Automa~ Front~ 4 2.3 Regu~ 27 19  
## 5 14039 Acura 2.3CL~ 1998 Subcom~ Manual~ Front~ 4 2.3 Regu~ 29 21  
## 6 14040 Acura 2.3CL~ 1998 Subcom~ Automa~ Front~ 6 3 Regu~ 26 17  
## 7 14834 Acura 2.3CL~ 1999 Subcom~ Automa~ Front~ 4 2.3 Regu~ 27 20  
## 8 14835 Acura 2.3CL~ 1999 Subcom~ Manual~ Front~ 4 2.3 Regu~ 29 21  
## 9 14836 Acura 2.3CL~ 1999 Subcom~ Automa~ Front~ 6 3 Regu~ 26 17  
## 10 16573 Acura 3.2CL 2001 Compac~ Automa~ Front~ 6 3.2 Prem~ 27 17  
## # ... with 7,079 more rows

filter(vehicles, between(displ, 1, 5))

## # A tibble: 28,686 x 12  
## id make model year class trans drive cyl displ fuel hwy cty  
## <int> <chr> <chr> <int> <chr> <chr> <chr> <int> <dbl> <chr> <int> <int>  
## 1 27550 AM Ge~ DJ Po ~ 1984 Specia~ Auto~ 2-Whe~ 4 2.5 Regu~ 17 18  
## 2 28426 AM Ge~ DJ Po ~ 1984 Specia~ Auto~ 2-Whe~ 4 2.5 Regu~ 17 18  
## 3 27549 AM Ge~ FJ8c P~ 1984 Specia~ Auto~ 2-Whe~ 6 4.2 Regu~ 13 13  
## 4 28425 AM Ge~ FJ8c P~ 1984 Specia~ Auto~ 2-Whe~ 6 4.2 Regu~ 13 13  
## 5 1032 AM Ge~ Post O~ 1985 Specia~ Auto~ Rear-~ 4 2.5 Regu~ 17 16  
## 6 1033 AM Ge~ Post O~ 1985 Specia~ Auto~ Rear-~ 6 4.2 Regu~ 13 13  
## 7 3347 ASC I~ GNX 1987 Midsiz~ Auto~ Rear-~ 6 3.8 Prem~ 21 14  
## 8 13309 Acura 2.2CL/~ 1997 Subcom~ Auto~ Front~ 4 2.2 Regu~ 26 20  
## 9 13310 Acura 2.2CL/~ 1997 Subcom~ Manu~ Front~ 4 2.2 Regu~ 28 22  
## 10 13311 Acura 2.2CL/~ 1997 Subcom~ Auto~ Front~ 6 3 Regu~ 26 18  
## # ... with 28,676 more rows

1. How many cars have a missing cyl? What other variables are often missing with cyl? What might these rows represent?

3A. There are 58 cars in this list that do not have cyl values. These are mainly electric vehicles that do not have traditional combustion chambers (cylinders). However, the Mazda RX-7 and the Ford Ranger Pickups from 1999-2001 are not electric and are just missing the cyl values.

filter(vehicles, is.na(cyl))

## # A tibble: 58 x 12  
## id make model year class trans drive cyl displ fuel hwy cty  
## <int> <chr> <chr> <int> <chr> <chr> <chr> <int> <dbl> <chr> <int> <int>  
## 1 31893 Azure~ Transi~ 2012 Specia~ Autom~ Fron~ NA NA Elec~ 62 62  
## 2 31894 Azure~ Transi~ 2012 Specia~ Autom~ Fron~ NA NA Elec~ 62 62  
## 3 32276 BMW Active~ 2011 Subcom~ Autom~ Rear~ NA NA Elec~ 96 107  
## 4 33383 BYD e6 2012 Sport ~ Autom~ Fron~ NA NA Elec~ 64 60  
## 5 34860 BYD e6 2013 Small ~ Autom~ Fron~ NA NA Elec~ 65 61  
## 6 34859 BYD e6 2014 Small ~ Autom~ Fron~ NA NA Elec~ 65 61  
## 7 32277 CODA ~ CODA 2012 Subcom~ Autom~ Fron~ NA NA Elec~ 68 77  
## 8 33337 CODA ~ CODA 2013 Subcom~ Autom~ Fron~ NA NA Elec~ 68 77  
## 9 30976 Chevr~ S10 El~ 1998 Small ~ Autom~ 2-Wh~ NA NA Elec~ 58 52  
## 10 30977 Chevr~ S10 El~ 1998 Small ~ Autom~ 2-Wh~ NA NA Elec~ 28 28  
## # ... with 48 more rows

1. Why is NA ^ 0 not missing? Why is NA | TRUE not missing? Why is FALSE & NA not missing? Can you figure out the general rule? (NA \* 0 is a tricky counterexample!)

4A. NA ^ 0 and NA | TRUE evaluate as 1 and TRUE respectively because they use an type of OR comparison operator, which ignores the presence of an unknown value. FALSE & NA evaluates as FALSE. I think the general rule is that when there is a comparison operator which allows us to ‘ignore’ the NA value.

1. How could you use arrange() to sort all missing values of cyl to the start? (Hint: use is.na()).

5A.

arrange(vehicles, !is.na(cyl), cyl)

## # A tibble: 33,442 x 12  
## id make model year class trans drive cyl displ fuel hwy cty  
## <int> <chr> <chr> <int> <chr> <chr> <chr> <int> <dbl> <chr> <int> <int>  
## 1 31893 Azure~ Transi~ 2012 Specia~ Autom~ Fron~ NA NA Elec~ 62 62  
## 2 31894 Azure~ Transi~ 2012 Specia~ Autom~ Fron~ NA NA Elec~ 62 62  
## 3 32276 BMW Active~ 2011 Subcom~ Autom~ Rear~ NA NA Elec~ 96 107  
## 4 33383 BYD e6 2012 Sport ~ Autom~ Fron~ NA NA Elec~ 64 60  
## 5 34860 BYD e6 2013 Small ~ Autom~ Fron~ NA NA Elec~ 65 61  
## 6 34859 BYD e6 2014 Small ~ Autom~ Fron~ NA NA Elec~ 65 61  
## 7 32277 CODA ~ CODA 2012 Subcom~ Autom~ Fron~ NA NA Elec~ 68 77  
## 8 33337 CODA ~ CODA 2013 Subcom~ Autom~ Fron~ NA NA Elec~ 68 77  
## 9 30976 Chevr~ S10 El~ 1998 Small ~ Autom~ 2-Wh~ NA NA Elec~ 58 52  
## 10 30977 Chevr~ S10 El~ 1998 Small ~ Autom~ 2-Wh~ NA NA Elec~ 28 28  
## # ... with 33,432 more rows

1. Sort vehicles to find the vehicles with the largest city miles per gallon. Also find the vehicles that had the lowest highway miles per gallon.

6A. Highest CTY mpg: 138 Lowest HWY mpg: 9

arrange(vehicles, desc(cty))

## # A tibble: 33,442 x 12  
## id make model year class trans drive cyl displ fuel hwy cty  
## <int> <chr> <chr> <int> <chr> <chr> <chr> <int> <dbl> <chr> <int> <int>  
## 1 33307 Scion iQ EV 2013 Minic~ Automa~ Front~ NA NA Elec~ 105 138  
## 2 33265 Honda Fit EV 2013 Small~ Automa~ Front~ NA NA Elec~ 105 132  
## 3 33905 Honda Fit EV 2014 Small~ Automa~ Front~ NA NA Elec~ 105 132  
## 4 33558 Nissan Leaf 2013 Midsi~ Automa~ Front~ NA NA Elec~ 102 129  
## 5 33640 Chevr~ Spark~ 2014 Subco~ Automa~ Front~ NA NA Elec~ 109 128  
## 6 31673 Mitsu~ i-MiEV 2012 Subco~ Automa~ Rear-~ NA NA Elec~ 99 126  
## 7 33358 Mitsu~ i-MiEV 2013 Subco~ Automa~ Rear-~ NA NA Elec~ 99 126  
## 8 34901 Mitsu~ i-MiEV 2014 Subco~ Automa~ Rear-~ NA NA Elec~ 99 126  
## 9 34699 Nissan Leaf 2014 Midsi~ Automa~ Front~ NA NA Elec~ 101 126  
## 10 34918 Nissan Leaf 2015 Midsi~ Automa~ Front~ NA NA Elec~ 101 126  
## # ... with 33,432 more rows

arrange(vehicles, hwy)

## # A tibble: 33,442 x 12  
## id make model year class trans drive cyl displ fuel hwy cty  
## <int> <chr> <chr> <int> <chr> <chr> <chr> <int> <dbl> <chr> <int> <int>  
## 1 1178 Aston~ Lagonda 1985 Subco~ Autom~ Rear-~ 8 5.3 Regu~ 9 7  
## 2 47 Aston~ Lagonda 1985 Subco~ Autom~ Rear-~ 8 5.3 Regu~ 9 7  
## 3 1179 Aston~ Saloon~ 1985 Subco~ Autom~ Rear-~ 8 5.3 Regu~ 9 7  
## 4 48 Aston~ Saloon~ 1985 Subco~ Autom~ Rear-~ 8 5.3 Regu~ 9 7  
## 5 3308 Rolls~ Camarg~ 1987 Compa~ Autom~ Rear-~ 8 6.8 Regu~ 9 7  
## 6 3110 Rolls~ Contin~ 1987 Subco~ Autom~ Rear-~ 8 6.8 Regu~ 9 7  
## 7 3113 Rolls~ Cornic~ 1987 Subco~ Autom~ Rear-~ 8 6.8 Regu~ 9 7  
## 8 3115 Rolls~ Cornic~ 1987 Subco~ Autom~ Rear-~ 8 6.8 Regu~ 9 7  
## 9 3444 Rolls~ Eight/~ 1987 Midsi~ Autom~ Rear-~ 8 6.8 Regu~ 9 8  
## 10 3447 Rolls~ Silver~ 1987 Midsi~ Autom~ Rear-~ 8 6.8 Regu~ 9 8  
## # ... with 33,432 more rows

1. Sort vehicles to find the vehicles with the largest number of cylnders.

7A. Largest number of cylinders: 16

arrange(vehicles, desc(cyl))

## # A tibble: 33,442 x 12  
## id make model year class trans drive cyl displ fuel hwy cty  
## <int> <chr> <chr> <int> <chr> <chr> <chr> <int> <dbl> <chr> <int> <int>  
## 1 22661 Bugat~ Veyron 2006 Two S~ Autom~ 4-Whee~ 16 8 Prem~ 14 8  
## 2 24169 Bugat~ Veyron 2008 Two S~ Autom~ 4-Whee~ 16 8 Prem~ 14 8  
## 3 29250 Bugat~ Veyron 2010 Two S~ Autom~ All-Wh~ 16 8 Prem~ 14 8  
## 4 30226 Bugat~ Veyron 2011 Two S~ Autom~ All-Wh~ 16 8 Prem~ 15 8  
## 5 31619 Bugat~ Veyron 2012 Two S~ Autom~ All-Wh~ 16 8 Prem~ 15 8  
## 6 32656 Bugat~ Veyron 2013 Two S~ Autom~ All-Wh~ 16 8 Prem~ 15 8  
## 7 33614 Bugat~ Veyron 2014 Two S~ Autom~ All-Wh~ 16 8 Prem~ 15 8  
## 8 19324 Aston~ DB AR1 2003 Two S~ Autom~ Rear-W~ 12 5.9 Prem~ 17 10  
## 9 19325 Aston~ DB AR1 2003 Two S~ Manua~ Rear-W~ 12 5.9 Prem~ 16 9  
## 10 18341 Aston~ DB-7 ~ 2003 Minic~ Manua~ Rear-W~ 12 5.9 Prem~ 15 9  
## # ... with 33,432 more rows

1. Which vehicles have the largest displacement? Which have the smallest?

8A. Largest displacement: 8.4 Smallest displacement: 0.0? or 1.0

arrange(vehicles, desc(displ))

## # A tibble: 33,442 x 12  
## id make model year class trans drive cyl displ fuel hwy cty  
## <int> <chr> <chr> <int> <chr> <chr> <chr> <int> <dbl> <chr> <int> <int>  
## 1 24752 Dodge Viper ~ 2008 Two Se~ Manua~ Rear-~ 10 8.4 Prem~ 22 13  
## 2 26295 Dodge Viper ~ 2009 Two Se~ Manua~ Rear-~ 10 8.4 Prem~ 22 13  
## 3 29895 Dodge Viper ~ 2010 Two Se~ Manua~ Rear-~ 10 8.4 Prem~ 22 13  
## 4 24753 Dodge Viper ~ 2008 Two Se~ Manua~ Rear-~ 10 8.4 Prem~ 22 13  
## 5 26296 Dodge Viper ~ 2009 Two Se~ Manua~ Rear-~ 10 8.4 Prem~ 22 13  
## 6 29896 Dodge Viper ~ 2010 Two Se~ Manua~ Rear-~ 10 8.4 Prem~ 22 13  
## 7 33534 SRT Viper 2013 Two Se~ Manua~ Rear-~ 10 8.4 Prem~ 19 12  
## 8 34131 SRT Viper 2014 Two Se~ Manua~ Rear-~ 10 8.4 Prem~ 19 12  
## 9 19979 Dodge Ram 15~ 2004 Standa~ Manua~ Rear-~ 10 8.3 Prem~ 14 9  
## 10 21105 Dodge Ram 15~ 2005 Standa~ Autom~ Rear-~ 10 8.3 Prem~ 11 8  
## # ... with 33,432 more rows

arrange(vehicles, displ)

## # A tibble: 33,442 x 12  
## id make model year class trans drive cyl displ fuel hwy cty  
## <int> <chr> <chr> <int> <chr> <chr> <chr> <int> <dbl> <chr> <int> <int>  
## 1 11257 GMC Sierr~ 1994 Standa~ Autom~ Rear-~ 8 0 Dies~ 20 15  
## 2 11258 GMC Sierr~ 1994 Standa~ Manua~ Rear-~ 8 0 Dies~ 22 17  
## 3 11341 GMC Sierr~ 1994 Standa~ Manua~ 4-Whe~ 8 0 Dies~ 19 15  
## 4 14064 Chevr~ Metro 1998 Subcom~ Manua~ Front~ 3 1 Regu~ 44 36  
## 5 14857 Chevr~ Metro 1999 Subcom~ Manua~ Front~ 3 1 Regu~ 42 34  
## 6 15659 Chevr~ Metro 2000 Subcom~ Manua~ Front~ 3 1 Regu~ 41 32  
## 7 36 Chevr~ Sprint 1985 Minico~ Autom~ Front~ 3 1 Regu~ 39 34  
## 8 37 Chevr~ Sprint 1985 Minico~ Manua~ Front~ 3 1 Regu~ 47 39  
## 9 1728 Chevr~ Sprint 1986 Minico~ Autom~ Front~ 3 1 Regu~ 38 34  
## 10 1729 Chevr~ Sprint 1986 Minico~ Manua~ Front~ 3 1 Regu~ 45 37  
## # ... with 33,432 more rows

1. Select all quantitative variables from vehicles.

9A. cyl, displ, hwy, cty

select(vehicles, cyl, displ, hwy, cty)

## # A tibble: 33,442 x 4  
## cyl displ hwy cty  
## <int> <dbl> <int> <int>  
## 1 4 2.5 17 18  
## 2 4 2.5 17 18  
## 3 6 4.2 13 13  
## 4 6 4.2 13 13  
## 5 4 2.5 17 16  
## 6 6 4.2 13 13  
## 7 6 3.8 21 14  
## 8 4 2.2 26 20  
## 9 4 2.2 28 22  
## 10 6 3 26 18  
## # ... with 33,432 more rows

1. What does the one\_of() function do? Rewrite the answer to the previous question using it. Hint: make sure the variable names are in "".

10A. The one\_of() function selects column names from the whole data set.

select(vehicles, one\_of("cyl", "displ", "hwy", "cty"))

## # A tibble: 33,442 x 4  
## cyl displ hwy cty  
## <int> <dbl> <int> <int>  
## 1 4 2.5 17 18  
## 2 4 2.5 17 18  
## 3 6 4.2 13 13  
## 4 6 4.2 13 13  
## 5 4 2.5 17 16  
## 6 6 4.2 13 13  
## 7 6 3.8 21 14  
## 8 4 2.2 26 20  
## 9 4 2.2 28 22  
## 10 6 3 26 18  
## # ... with 33,432 more rows

1. What happens if you include the name of a variable multiple times in a select() call?

11A. dplyr will ignore a repeated variable parameter and only show the variable in the first position it was called (left to right)

select(vehicles, cty, cyl, displ, hwy, cty, cty)

## # A tibble: 33,442 x 4  
## cty cyl displ hwy  
## <int> <int> <dbl> <int>  
## 1 18 4 2.5 17  
## 2 18 4 2.5 17  
## 3 13 6 4.2 13  
## 4 13 6 4.2 13  
## 5 16 4 2.5 17  
## 6 13 6 4.2 13  
## 7 14 6 3.8 21  
## 8 20 4 2.2 26  
## 9 22 4 2.2 28  
## 10 18 6 3 26  
## # ... with 33,432 more rows

1. Make a new variable hwygpm where hwygpm=gallons per mile on the highway=1/hwy

12A.

vehicles2 = mutate(vehicles, hwygpm = 1/hwy)  
vehicles2

## # A tibble: 33,442 x 13  
## id make model year class trans drive cyl displ fuel hwy cty  
## <int> <chr> <chr> <int> <chr> <chr> <chr> <int> <dbl> <chr> <int> <int>  
## 1 27550 AM G~ DJ P~ 1984 Spec~ Auto~ 2-Wh~ 4 2.5 Regu~ 17 18  
## 2 28426 AM G~ DJ P~ 1984 Spec~ Auto~ 2-Wh~ 4 2.5 Regu~ 17 18  
## 3 27549 AM G~ FJ8c~ 1984 Spec~ Auto~ 2-Wh~ 6 4.2 Regu~ 13 13  
## 4 28425 AM G~ FJ8c~ 1984 Spec~ Auto~ 2-Wh~ 6 4.2 Regu~ 13 13  
## 5 1032 AM G~ Post~ 1985 Spec~ Auto~ Rear~ 4 2.5 Regu~ 17 16  
## 6 1033 AM G~ Post~ 1985 Spec~ Auto~ Rear~ 6 4.2 Regu~ 13 13  
## 7 3347 ASC ~ GNX 1987 Mids~ Auto~ Rear~ 6 3.8 Prem~ 21 14  
## 8 13309 Acura 2.2C~ 1997 Subc~ Auto~ Fron~ 4 2.2 Regu~ 26 20  
## 9 13310 Acura 2.2C~ 1997 Subc~ Manu~ Fron~ 4 2.2 Regu~ 28 22  
## 10 13311 Acura 2.2C~ 1997 Subc~ Auto~ Fron~ 6 3 Regu~ 26 18  
## # ... with 33,432 more rows, and 1 more variable: hwygpm <dbl>

1. Make a new variable ctykpl where ctykpl=kilometers per liter in the city=0.42517\*cty

13A.

vehicles2 = mutate(vehicles2, ctykpl = 0.42517\*cty)  
vehicles2

## # A tibble: 33,442 x 14  
## id make model year class trans drive cyl displ fuel hwy cty  
## <int> <chr> <chr> <int> <chr> <chr> <chr> <int> <dbl> <chr> <int> <int>  
## 1 27550 AM G~ DJ P~ 1984 Spec~ Auto~ 2-Wh~ 4 2.5 Regu~ 17 18  
## 2 28426 AM G~ DJ P~ 1984 Spec~ Auto~ 2-Wh~ 4 2.5 Regu~ 17 18  
## 3 27549 AM G~ FJ8c~ 1984 Spec~ Auto~ 2-Wh~ 6 4.2 Regu~ 13 13  
## 4 28425 AM G~ FJ8c~ 1984 Spec~ Auto~ 2-Wh~ 6 4.2 Regu~ 13 13  
## 5 1032 AM G~ Post~ 1985 Spec~ Auto~ Rear~ 4 2.5 Regu~ 17 16  
## 6 1033 AM G~ Post~ 1985 Spec~ Auto~ Rear~ 6 4.2 Regu~ 13 13  
## 7 3347 ASC ~ GNX 1987 Mids~ Auto~ Rear~ 6 3.8 Prem~ 21 14  
## 8 13309 Acura 2.2C~ 1997 Subc~ Auto~ Fron~ 4 2.2 Regu~ 26 20  
## 9 13310 Acura 2.2C~ 1997 Subc~ Manu~ Fron~ 4 2.2 Regu~ 28 22  
## 10 13311 Acura 2.2C~ 1997 Subc~ Auto~ Fron~ 6 3 Regu~ 26 18  
## # ... with 33,432 more rows, and 2 more variables: hwygpm <dbl>, ctykpl <dbl>

1. Make a new variable that is the dfference between hwy and cty.

14A.

vehicles2 = mutate(vehicles2, mpgDif = hwy-cty)  
vehicles2

## # A tibble: 33,442 x 15  
## id make model year class trans drive cyl displ fuel hwy cty  
## <int> <chr> <chr> <int> <chr> <chr> <chr> <int> <dbl> <chr> <int> <int>  
## 1 27550 AM G~ DJ P~ 1984 Spec~ Auto~ 2-Wh~ 4 2.5 Regu~ 17 18  
## 2 28426 AM G~ DJ P~ 1984 Spec~ Auto~ 2-Wh~ 4 2.5 Regu~ 17 18  
## 3 27549 AM G~ FJ8c~ 1984 Spec~ Auto~ 2-Wh~ 6 4.2 Regu~ 13 13  
## 4 28425 AM G~ FJ8c~ 1984 Spec~ Auto~ 2-Wh~ 6 4.2 Regu~ 13 13  
## 5 1032 AM G~ Post~ 1985 Spec~ Auto~ Rear~ 4 2.5 Regu~ 17 16  
## 6 1033 AM G~ Post~ 1985 Spec~ Auto~ Rear~ 6 4.2 Regu~ 13 13  
## 7 3347 ASC ~ GNX 1987 Mids~ Auto~ Rear~ 6 3.8 Prem~ 21 14  
## 8 13309 Acura 2.2C~ 1997 Subc~ Auto~ Fron~ 4 2.2 Regu~ 26 20  
## 9 13310 Acura 2.2C~ 1997 Subc~ Manu~ Fron~ 4 2.2 Regu~ 28 22  
## 10 13311 Acura 2.2C~ 1997 Subc~ Auto~ Fron~ 6 3 Regu~ 26 18  
## # ... with 33,432 more rows, and 3 more variables: hwygpm <dbl>, ctykpl <dbl>,  
## # mpgDif <int>

1. How many vehicles have a negative difference between hwy and cty? Any common features of some of the vehicles?

15A. Vehicles with negative hwy/cty differentials tend to be electric vehicles which are less efficient in the stop-and-go of cities.

arrange(vehicles2, mpgDif)

## # A tibble: 33,442 x 15  
## id make model year class trans drive cyl displ fuel hwy cty  
## <int> <chr> <chr> <int> <chr> <chr> <chr> <int> <dbl> <chr> <int> <int>  
## 1 33307 Scion iQ EV 2013 Mini~ Auto~ Fron~ NA NA Elec~ 105 138  
## 2 33305 smart fort~ 2013 Two ~ Auto~ Rear~ NA NA Elec~ 93 122  
## 3 34393 smart fort~ 2014 Two ~ Auto~ Rear~ NA NA Elec~ 93 122  
## 4 33306 smart fort~ 2013 Two ~ Auto~ Rear~ NA NA Elec~ 93 122  
## 5 34394 smart fort~ 2014 Two ~ Auto~ Rear~ NA NA Elec~ 93 122  
## 6 33265 Honda Fit ~ 2013 Smal~ Auto~ Fron~ NA NA Elec~ 105 132  
## 7 33905 Honda Fit ~ 2014 Smal~ Auto~ Fron~ NA NA Elec~ 105 132  
## 8 31673 Mits~ i-Mi~ 2012 Subc~ Auto~ Rear~ NA NA Elec~ 99 126  
## 9 33358 Mits~ i-Mi~ 2013 Subc~ Auto~ Rear~ NA NA Elec~ 99 126  
## 10 34901 Mits~ i-Mi~ 2014 Subc~ Auto~ Rear~ NA NA Elec~ 99 126  
## # ... with 33,432 more rows, and 3 more variables: hwygpm <dbl>, ctykpl <dbl>,  
## # mpgDif <int>