dyplr

```
library(tidyverse) # importing tidyverse for dplyr
```

```
## -- Attaching packages --------------------------------------- tidyverse
1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.8
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
```

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

filter

```
head(airquality %>% filter(Day > 5, Temp > 70))
```

```
##   Ozone Solar.R Wind Temp Month Day
## 1     7      NA  6.9   74     5  11
## 2    11     320 16.6   73     5  22
## 3    45     252 14.9   81     5  29
## 4   115     223  5.7   79     5  30
## 5    37     279  7.4   76     5  31
## 6    NA     264 14.3   79     6   6
```

```
airquality %>% filter(Temp == 70)
```

```
##   Ozone Solar.R Wind Temp Month Day
## 1    30     193  6.9   70     9  26
```

arrange

```
head(airquality %>% arrange(Month, Day, Wind))
```

```
##   Ozone Solar.R Wind Temp Month Day
## 1    41     190  7.4   67     5   1
## 2    36     118  8.0   72     5   2
## 3    12     149 12.6   74     5   3
## 4    18     313 11.5   62     5   4
## 5    NA      NA 14.3   56     5   5
## 6    28      NA 14.9   66     5   6
```

```
head(airquality %>% arrange(desc(Month), Day, Wind) )
```

```
##   Ozone Solar.R Wind Temp Month Day
## 1    96     167  6.9   91     9   1
## 2    78     197  5.1   92     9   2
## 3    73     183  2.8   93     9   3
## 4    91     189  4.6   93     9   4
```

```
## 5    47      95  7.4   87     9   5
## 6    32      92 15.5   84     9   6
```

slice

```
head(airquality %>% slice(5:10))

##   Ozone Solar.R Wind Temp Month Day
## 1    NA      NA 14.3   56     5   5
## 2    28      NA 14.9   66     5   6
## 3    23     299  8.6   65     5   7
## 4    19      99 13.8   59     5   8
## 5     8      19 20.1   61     5   9
## 6    NA     194  8.6   69     5  10

airquality %>% slice(1:5)

##   Ozone Solar.R Wind Temp Month Day
## 1    41     190  7.4   67     5   1
## 2    36     118  8.0   72     5   2
## 3    12     149 12.6   74     5   3
## 4    18     313 11.5   62     5   4
## 5    NA      NA 14.3   56     5   5

airquality %>% slice_head(n=3)

##   Ozone Solar.R Wind Temp Month Day
## 1    41     190  7.4   67     5   1
## 2    36     118  8.0   72     5   2
## 3    12     149 12.6   74     5   3

airquality %>% slice_tail(n=3)

##   Ozone Solar.R Wind Temp Month Day
## 1    14     191 14.3   75     9  28
## 2    18     131  8.0   76     9  29
## 3    20     223 11.5   68     9  30

head(airquality %>% sample(2))

##   Ozone Day
## 1    41   1
## 2    36   2
## 3    12   3
## 4    18   4
## 5    NA   5
## 6    28   6

airquality %>% sample_n(2)

##   Ozone Solar.R Wind Temp Month Day
## 1    NA     139  8.6   82     7  11
## 2    NA     186  9.2   84     6   4
```

```
airquality %>% slice_sample(n=5)
```

```
##   Ozone Solar.R Wind Temp Month Day
## 1    14     274 10.9   68     5  14
## 2    NA     194  8.6   69     5  10
## 3    NA     250  6.3   76     6  24
## 4   108     223  8.0   85     7  25
## 5    NA      31 14.9   77     6  29
```

```
head(airquality %>% slice_sample(prop = .2)) #20 percent of data is sampled
```

```
##   Ozone Solar.R Wind Temp Month Day
## 1    NA     127  8.0   78     6  26
## 2    12     149 12.6   74     5   3
## 3    84     237  6.3   96     8  30
## 4    32      92 12.0   61     5  24
## 5    16     201  8.0   82     9  20
## 6    61     285  6.3   84     7  18
```

select

```
head(select(airquality, Ozone, Wind))
```

```
##   Ozone Wind
## 1    41  7.4
## 2    36  8.0
## 3    12 12.6
## 4    18 11.5
## 5    NA 14.3
## 6    28 14.9
```

```
head(select(airquality, 2, 4))
```

```
##   Solar.R Temp
## 1     190   67
## 2     118   72
## 3     149   74
## 4     313   62
## 5      NA   56
## 6      NA   66
```

```
head(airquality %>% select(Ozone, Wind))
```

```
##   Ozone Wind
## 1    41  7.4
## 2    36  8.0
## 3    12 12.6
## 4    18 11.5
## 5    NA 14.3
## 6    28 14.9
```

```
head(airquality %>% select(Ozone:Wind))
```

```
##   Ozone Solar.R Wind
## 1    41     190  7.4
## 2    36     118  8.0
## 3    12     149 12.6
## 4    18     313 11.5
## 5    NA      NA 14.3
## 6    28      NA 14.9
```

```
head(airquality %>% select(!Ozone))
```

```
##   Solar.R Wind Temp Month Day
## 1     190  7.4   67     5   1
## 2     118  8.0   72     5   2
## 3     149 12.6   74     5   3
## 4     313 11.5   62     5   4
## 5      NA 14.3   56     5   5
## 6      NA 14.9   66     5   6
```

```
head(airquality %>% select(!Ozone:Wind))
```

```
##   Temp Month Day
## 1   67     5   1
## 2   72     5   2
## 3   74     5   3
## 4   62     5   4
## 5   56     5   5
## 6   66     5   6
```

```
head(airquality %>% select(ends_with('nth')))
```

```
##   Month
## 1     5
## 2     5
## 3     5
## 4     5
## 5     5
## 6     5
```

```
head(airquality %>% select(starts_with('So')))
```

```
##   Solar.R
## 1     190
## 2     118
## 3     149
## 4     313
## 5      NA
## 6      NA
```

```
head(airquality %>% select(contains('ol')))
```

```
##   Solar.R
## 1     190
```

```
## 2     118
## 3     149
## 4     313
## 5      NA
## 6      NA
```

mutate

```
head(mutate(airquality, newWind = Wind + 10))
```

```
##   Ozone Solar.R Wind Temp Month Day newWind
## 1    41     190  7.4   67     5   1    17.4
## 2    36     118  8.0   72     5   2    18.0
## 3    12     149 12.6   74     5   3    22.6
## 4    18     313 11.5   62     5   4    21.5
## 5    NA      NA 14.3   56     5   5    24.3
## 6    28      NA 14.9   66     5   6    24.9
```

```
head(airquality %>% mutate(Wind = Wind * 10))
```

```
##   Ozone Solar.R Wind Temp Month Day
## 1    41     190   74   67     5   1
## 2    36     118   80   72     5   2
## 3    12     149  126   74     5   3
## 4    18     313  115   62     5   4
## 5    NA      NA  143   56     5   5
## 6    28      NA  149   66     5   6
```

```
head(airquality %>% mutate(Wind = ifelse(is.na(Solar.R), mean(Solar.R, na.rm
= T), Solar.R)))
```

```
##   Ozone Solar.R     Wind Temp Month Day
## 1    41     190 190.0000   67     5   1
## 2    36     118 118.0000   72     5   2
## 3    12     149 149.0000   74     5   3
## 4    18     313 313.0000   62     5   4
## 5    NA      NA 185.9315   56     5   5
## 6    28      NA 185.9315   66     5   6
```

```
head(airquality %>% mutate(Wind / Temp))
```

```
##   Ozone Solar.R Wind Temp Month Day Wind/Temp
## 1    41     190  7.4   67     5   1 0.1104478
## 2    36     118  8.0   72     5   2 0.1111111
## 3    12     149 12.6   74     5   3 0.1702703
## 4    18     313 11.5   62     5   4 0.1854839
## 5    NA      NA 14.3   56     5   5 0.2553571
## 6    28      NA 14.9   66     5   6 0.2257576
```

```
head(airquality %>% transmute(Wind / Temp))
```

```
##   Wind/Temp
## 1 0.1104478
```

```
## 2 0.1111111
## 3 0.1702703
## 4 0.1854839
## 5 0.2553571
## 6 0.2257576
```

relocate

```
head(airquality %>% relocate(Wind, .after = Month))
```

```
##   Ozone Solar.R Temp Month Wind Day
## 1    41     190   67     5  7.4   1
## 2    36     118   72     5  8.0   2
## 3    12     149   74     5 12.6   3
## 4    18     313   62     5 11.5   4
## 5    NA      NA   56     5 14.3   5
## 6    28      NA   66     5 14.9   6
```

```
head(airquality %>% relocate(Solar.R:Temp, .after = Month))
```

```
##   Ozone Month Solar.R Wind Temp Day
## 1    41     5     190  7.4   67   1
## 2    36     5     118  8.0   72   2
## 3    12     5     149 12.6   74   3
## 4    18     5     313 11.5   62   4
## 5    NA     5      NA 14.3   56   5
## 6    28     5      NA 14.9   66   6
```

summarise

```
airquality %>% summarise(mean(Month),sd(Month), sum(Month), min(Month),
max(Month))
```

```
##   mean(Month) sd(Month) sum(Month) min(Month) max(Month)
## 1    6.993464  1.416522       1070          5          9
```

```
airquality %>% summarise(var(Month))
```

```
##   var(Month)
## 1   2.006536
```

across

```
summarise(airquality, across(everything(), mean))
```

```
##   Ozone Solar.R     Wind     Temp    Month      Day
## 1    NA      NA 9.957516 77.88235 6.993464 15.80392
```

groupby

```
airquality %>% group_by(Month) %>% select(Wind) %>% summarise(mean(Wind))
```

```
## Adding missing grouping variables: `Month`
```

```
## # A tibble: 5 x 2
##   Month `mean(Wind)`
```

```
##   <int>        <dbl>
## 1     5        11.6
## 2     6        10.3
## 3     7         8.94
## 4     8         8.79
## 5     9        10.2
```

```
airquality %>% group_by(Month) %>% select(Wind, Temp, Ozone) %>%
summarise(mean(Wind), min(Temp), max(Ozone))
```

```
## Adding missing grouping variables: `Month`
```

```
## # A tibble: 5 x 4
##   Month `mean(Wind)` `min(Temp)` `max(Ozone)`
##   <int>        <dbl>       <int>        <int>
## 1     5        11.6           56           NA
## 2     6        10.3           65           NA
## 3     7         8.94          73           NA
## 4     8         8.79          72           NA
## 5     9        10.2           63           NA
```

count

```
count(airquality, Month)
```

```
##   Month  n
## 1     5 31
## 2     6 30
## 3     7 31
## 4     8 31
## 5     9 30
```

distinct

```
distinct(airquality, Month)
```

```
##   Month
## 1     5
## 2     6
## 3     7
## 4     8
## 5     9
```

pull

```
pull(airquality, Month)
```

```
##   [1] 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 6 6 6
6 6 6
##  [38] 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 7 7 7 7 7 7 7 7 7
7 7 7
##  [75] 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8
8 8 8
## [112] 8 8 8 8 8 8 8 8 8 8 8 8 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9
```

```
9 9 9
## [149] 9 9 9 9 9
```

rename

```
head(rename(airquality, breeze = Wind))
```

```
##   Ozone Solar.R breeze Temp Month Day
## 1    41     190    7.4   67     5   1
## 2    36     118    8.0   72     5   2
## 3    12     149   12.6   74     5   3
## 4    18     313   11.5   62     5   4
## 5    NA      NA   14.3   56     5   5
## 6    28      NA   14.9   66     5   6
```

```
head(rename(airquality, breeze = Wind, Temperature = Temp))
```

```
##   Ozone Solar.R breeze Temperature Month Day
## 1    41     190    7.4          67     5   1
## 2    36     118    8.0          72     5   2
## 3    12     149   12.6          74     5   3
## 4    18     313   11.5          62     5   4
## 5    NA      NA   14.3          56     5   5
## 6    28      NA   14.9          66     5   6
```

```
n_distinct(airquality)
```

```
## [1] 153
```

```
first(airquality)
```

```
##   [1]  41  36  12  18  NA  28  23  19   8  NA   7  16  11  14  18  14  34
6
##  [19]  30  11   1  11   4  32  NA  NA  NA  23  45 115  37  NA  NA  NA  NA
NA
##  [37]  NA  29  NA  71  39  NA  NA  23  NA  NA  21  37  20  12  13  NA  NA
NA
##  [55]  NA  NA  NA  NA  NA  NA  NA 135  49  32  NA  64  40  77  97  97  85
NA
##  [73]  10  27  NA   7  48  35  61  79  63  16  NA  NA  80 108  20  52  82
50
##  [91]  64  59  39   9  16  78  35  66 122  89 110  NA  NA  44  28  65  NA
22
## [109]  59  23  31  44  21   9  NA  45 168  73  NA  76 118  84  85  96  78
73
## [127]  91  47  32  20  23  21  24  44  21  28   9  13  46  18  13  24  16
13
## [145]  23  36   7  14  30  NA  14  18  20
```