WS #5 - Verbs

Wednesday, February 5, 2025

DS 002R - Jo Hardin

Name: _____

Names of people you worked with: _____

Work in groups of 3-4. Do you remember everyone's name? Tell your group about one talk/performance/event/activity not related to your classes that you are looking forward to in the coming weeks.

Task: Consider the `diamonds` dataset (all the variables names are given). Below are 2 tasks which can be accomplished using the following syntax. **Identify the data verbs and arguments for accomplishing each task** (the dataset includes the columns `x`, `y`, and `z` which are length, width, and depth in mm). Note, you may not need the last `arrange()`, but it won't cause errors.[1]

```
diamonds |>
  verb1( args1 ) |>
  verb2( args2 ) |>
  arrange( args3 ) |>
  head(1)
```

```
# A tibble: 3 x 10
  carat cut     color clarity depth table price     x     y     z
  <dbl> <ord>   <ord> <ord>   <dbl> <dbl> <int> <dbl> <dbl> <dbl>
1  0.23 Ideal   E     SI2      61.5    55   326  3.95  3.98  2.43
2  0.21 Premium E     SI1      59.8    61   326  3.89  3.84  2.31
3  0.23 Good    E     VS1      56.9    65   327  4.05  4.07  2.31
```

1. Which color diamond in this dataset is the largest on average (in terms of carats)?
2. What is the average price per carat of diamonds for the subset of diamonds that cost more than $10,000 total?

[1] From **Data Computing**, Daniel Kaplan

Solution:

1. Which color diamond in this dataset is the largest on average (in terms of carats)?

```
diamonds |>
  group_by( color ) |>
  summarize( avesize = mean(carat) ) |>
  arrange( desc(avesize) ) |>
  head(1)
```

```
# A tibble: 1 x 2
  color avesize
  <ord>   <dbl>
1 J        1.16
```

2. What is the average price per carat of diamonds for the subset of diamonds that cost more than $10,000 total?

```
diamonds |>
  filter(price > 10000) |>
  summarise( mean.ppc = mean(price/carat) ) |>
  arrange( desc(mean.ppc) ) |>
  head(1)
```

```
# A tibble: 1 x 1
  mean.ppc
     <dbl>
1    8044.
```