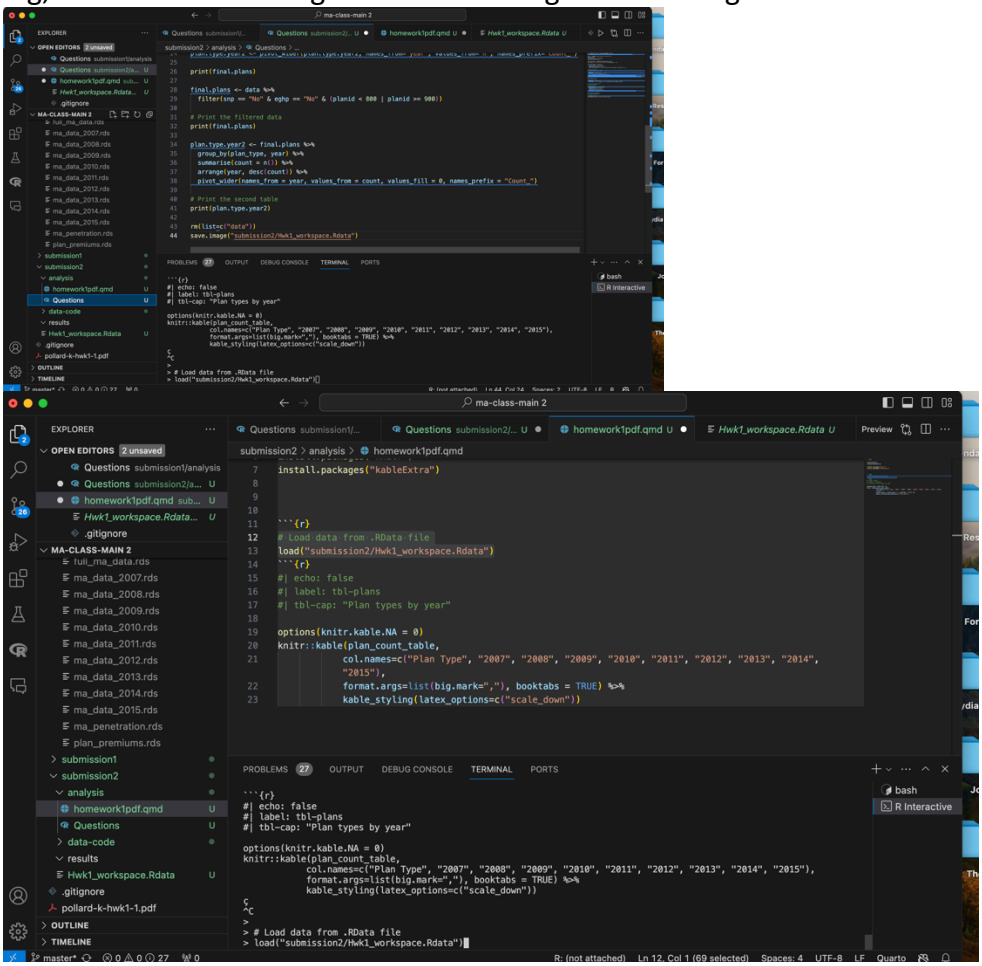
Kendall Pollard
HW1 Attempt 2

This attempt took a lot of time for me in trying to convert it to a PDF from Quarto. I ended up not removing (with the rm function) some of the read rds stuff of the huge files, so it brought that over to my HW1_workspace. When I pushed changes to GitHub, it broke. The files were too big, and then I couldn't go back and unstage those changes. So I had to start completely over.



Here is my new repository: https://github.com/kpollard8/Homework1. (that period is included)

Note: I didn't have question 4 in my initial submission done, but in this one it looks like I did. Sorry about this – I copied and pasted my code from where I currently was today, not where I was before class on Monday.

So for Submission 2:

- I worked for a long time on trying to get Quarto to then create a PDF and remove everything that would cause GitHub breaking issue
- I opened the Quarto file and started trying to make a pretty-looking table, but I couldn't get past the first loading

- Finished question 4,
- started on 5 and a little of 6.

Here are my answers to most of the questions so far:
Questions 1 – 3

1. How many observations exist in your current dataset?
    a. 19126783
2. How many different *plan_types* exist in the data?
    a. 27
3. Provide a table of the count of plans under each plan type in each year.

```
[1] 19126783
Number of different plan types 27
`summarise()` has grouped output by 'year'. You can override using the `.groups` argument.
# A tibble: 27 × 10
   plan_type      `2007` `2008` `2009` `2010` `2011` `2012` `2013` `2014` `2015`
   <chr>           <int>  <int>  <int>  <int>  <int>  <int>  <int>  <int>  <int>
 1 1876 Cost        5855   5459   5825   6035   6851   7633   7731 7.07e3   7157
 2 Continuing Ca…     95    122    158    142      0      0      0 0           0
 3 ESRD I             75    122    123    117      0      0      0 0           0
 4 ESRD II            12     12      7      8      0      0      0 0           0
 5 Employer Dire…   3247      0      0      0      0      0      0 0           0
 6 Employer/Unio…  32358  29113  25860  28700  28697  28669  25526 2.55e4  25630
 7 HCPP - 1833 C…     13     13   3938   3604     11     11     10 9   e0      9
 8 HMO/HMOPOS      60012  70176 479978 506802 528473 507272 530909 5.23e5 479275
 9 Local PPO       17427  38470 405197 417551 515700 636701 633884 6.65e5 704993
10 MA Health Sen…     73      0      0      0      0      0      0 0           0
11 MN Disability…     21      0      0      0      0      0      0 0           0
12 MN Senior Hea…    968      0      0      0      0      0      0 0           0
13 MSA               4422  16515  12267    135   6421   6416   6431 6.45e3   6518
14 MSA Demo          3274      0      0      0      0      0      0 0           0
15 Medicare Pres… 920058 963478 945794 893609 771694 815223 826907 1.12e6 991457
16 National PACE     405    548    616    717    781    858    953 1.12e3   1216
17 PFFS           364285 630756 683361 385733  45781  36423  31919 2.49e4  13658
18 PSO (Federal …    162      0      0      0      0      0      0 0           0
19 PSO (State Li…    421    535     87    123    176    171      0 0           0
20 Pilot              15     12    201     53      3      3      2 2   e0      2
21 Regional PPO    26402  27990  25943  24442  22773  21602  19970 1.98e4  17578
22 SHMO             1125      0      0      0      0      0      0 0           0
23 WI Partnershi…     42      0      0      0      0      0      0 0           0
24 Employer/Unio…      0   3332   3335   3332   3329   3323      0 0           0
25 RFB PFFS            0      0   3006      0      0      0      0 0           0
26 NA                  0      0  27505 277533      0      0      0 0           0
27 Medicare-Medi…      0      0      0      0      0      0    265 1.32e3   4130
```

```
> plan.type.year2 <- final.plans %>% group_by(plan_type, year) %>% count () %>% arrange(year, -n)
> plan.type.year2 <- pivot_wider(plan.type.year2, names_from="year", values_from="n", names_prefix="Count_")
> print(final.plans)
# A tibble: 4,074,476 × 18
# Groups:   contractid, planid, fips [1,424,607]
   contractid planid org_type  plan_type partd snp   eghp  org_name
   <chr>       <dbl> <chr>     <chr>     <chr> <chr> <chr> <chr>
 1 H0104           2 Local CCP Local PPO Yes   No    No    BLUE CROSS AND BLUE …
 2 H0104           2 Local CCP Local PPO Yes   No    No    BLUE CROSS AND BLUE …
 3 H0104           2 Local CCP Local PPO Yes   No    No    BLUE CROSS AND BLUE …
 4 H0104           2 Local CCP Local PPO Yes   No    No    BLUE CROSS AND BLUE …
 5 H0104           2 Local CCP Local PPO Yes   No    No    BLUE CROSS AND BLUE …
 6 H0104           2 Local CCP Local PPO Yes   No    No    BLUE CROSS AND BLUE …
 7 H0104           2 Local CCP Local PPO Yes   No    No    BLUE CROSS AND BLUE …
 8 H0104           2 Local CCP Local PPO Yes   No    No    BLUE CROSS AND BLUE …
 9 H0104           2 Local CCP Local PPO Yes   No    No    BLUE CROSS AND BLUE …
10 H0104           2 Local CCP Local PPO Yes   No    No    BLUE CROSS AND BLUE …
# ℹ 4,074,466 more rows
# ℹ 10 more variables: org_marketing_name <chr>, plan_name <chr>,
#   parent_org <chr>, contract_date <chr>, ssa <dbl>, fips <dbl>, state <chr>,
#   county <chr>, avg_enrollment <dbl>, year <int>
# ℹ Use `print(n = ...)` to see more rows
```

4.
5. Can merge with inner_join
6. Again use the inner_join
7. Couldn't do because I couldn't do 6
8. Why did we drop the "800-series" plans?
    a. I think 800 series plans are extensions of eghps. According to MA, employers engage with a third party, which assumes the insurance risk and operates as the plan sponsor. The designation "800 series" commonly encompasses the majority of EGWPs. Therefore, it's similar to a eghp, so we dropped it because we were already dropping the rest.
9. Why do so many plans charge a $0 premium? What does that really mean to a beneficiary?
    a. A $0 premium is attractive to beneficiaries because that means they don't have to pay an additional premium beyond their standard Medicare Part B premium. So I think it means that they still have to pay a premium, just not an additional
10. Briefly describe your experience working with these data (just a few sentences). Tell me one thing you learned and one thing that really aggravated you.
    a. Figuring out how to link it to a GitHub repository was incredibly exhausting – I had to start over 8x. The .gitignore wasn't working, so it would try to push all of the data files along with the code. I met with people in the class 4x this weekend. The thing that took the longest was setting up how to even start answering the questions. I ran out of time for 5, 6, 7 for that reason.
    b. I learned how to connect it to GitHub eventually and ignore the data files after multiple trial and error
    c. In the second attempt, I learned how to use a .qdm file, I learned how to create a "_workspace" file that only has my answers, and I started trying to make it a PDF – this was really time consuming