

Module 0: Getting Started

Part 3: Basics of Data Management in R

Ian McCarthy | Emory University
Econ 470 & HLTH 470

Real World

Practice data versus the real world



Advice 1: Be patient and careful in your coding



Advice 2: Comment, comment, comment

You don't want to end up like this guy...



Medicare Advantage

Medicare Advantage

Let's work with the [Medicare Advantage GitHub repository](#)

Access the data

First step is to download the raw data that we'll be using, or work in AWS:

- Monthly Enrollment
- Plan Characteristics
- Service Areas

Lots more out there, but this is enough for now.

1) Contract/enrollment info

```
for (y in 2007:2015) {  
  ## Basic contract/plan information  
  ma.path=paste0(path.data.ma, "/monthly-ma-and-pdp-enrollment-by-cpsc/CPSC_Contract_Info_")  
  contract.info=read_csv(ma.path,  
                          skip=1,  
                          col_names = c("contractid", "planid", "org_type", "plan_type",  
                                         "partd", "snp", "eghp", "org_name", "org_marketing_name",  
                                         "plan_name", "parent_org", "contract_date"),  
                          col_types = cols(  
                            contractid = col_character(),  
                            planid = col_double(),  
                            ...  
                          ))  
}
```

1) Contract/enrollment info

```
## Clean the contract level data  
contract.info = contract.info %>%  
  group_by(contractid, planid) %>%  
  mutate(id_count=row_number())  
  
contract.info = contract.info %>%  
  group_by(contractid, planid) %>%  
  mutate(id_count=row_number())
```

1) Contract/enrollment info

```
## Enrollments per plan
enroll.info=read_csv(paste0("data/input/monthly-ma-and-pdp-enrollment-by-cpsc/CPSC_Enrol
    skip=1,
    col_names = c("contractid","planid","ssa","fips","state","county","
    col_types = cols(
    contractid = col_character(),
    planid = col_double(),
    ssa = col_double(),
    fips = col_double(),
    state = col_character(),
    county = col_character(),
    enrollment = col_double()
    ),na="✖")
```

1) Contract/enrollment info

```
## Merge contract info with enrollment info  
plan.data = contract.info %>%  
  left_join(enroll.info, by=c("contractid", "planid")) %>%  
  mutate(year=y)
```

1) Contract/enrollment info

```
## Fill in missing fips codes (by state and county)
```

```
plan.data = plan.data %>%  
  group_by(state, county) %>%  
  fill(fips)
```

```
## Fill in missing plan characteristics by contract and plan id
```

```
plan.data = plan.data %>%  
  group_by(contractid, planid) %>%  
  fill(plan_type, partd, snp, eg hp, plan_name)
```

```
## Fill in missing contract characteristics by contractid
```

```
plan.data = plan.data %>%  
  group_by(contractid) %>%  
  fill(org_type, org_name, org_marketing_name, parent_org)
```

1) Contract/enrollment info

```
## Collapse from monthly data to yearly  
plan.year = plan.data %>%  
  group_by(contractid, planid, fips) %>%  
  arrange(contractid, planid, fips) %>%  
  rename(avg_enrollment=enrollment)  
  
write_rds(plan.year,paste0("data/output/ma_data_",y,".rds"))
```

1) Contract/enrollment info

```
full.ma.data ← read_rds("data/output/ma_data_2007.rds")
for (y in 2008:2015) {
  full.ma.data ← rbind(full.ma.data, read_rds(paste0("data/output/ma_data_", y, ".rds")))
}
```

All together now

Now let's do this together...

1. Copy data from OneDrive
2. Initialize repository in GitHub
3. Clone to local computer
4. Follow some practice code, available [here](#)