

Last login: Mon Feb 26 18:40:19 on ttys001

The default interactive shell is now zsh.
To update your account to use zsh, please run `chsh -s /bin/zsh`.
For more details, please visit <https://support.apple.com/kb/HT208050>.
(base) a-10-27-23-46:~ zoehaskellcraig\$ gcloud auth login --no-launch-browser

Go to the following link in your browser:

https://accounts.google.com/o/oauth2/auth?response_type=code&client_id=32555940559.apps.googleusercontent.com&redirect_uri=https%3A%2F%2Fsdk.cloud.google.com%2Fauthcode.html&scope=openid+https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fuserinfo.email+https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fcloud-platform+https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fappengine.admin+https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fsqlservice.login+https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fcompute+https%3A%2F%2Fwww.googleapis.com%2Fauth%2Faccounts.reauth&state=IoH0n2mUBESdf81FKIoCpjd0es14vv&prompt=consent&access_type=offline&code_challenge=j2uc15Y9d54iqQ6TKLXXIp0vyRJ6UXWZsuqR5jgasI&code_challenge_method=S256

Enter authorization code: **ERROR:** There was a problem with web authentication.
ERROR: (gcloud.auth.login) Please supply either code or authorization_response parameters.
(base) a-10-27-23-46:~ zoehaskellcraig\$ 4/0AeaYSHAB-tu1utUER_1Y1aR1f5E5UIMWIptdA-LZLHeE4DPqdgJeesDP75DF2z2Jclltjg
-bash: 4/0AeaYSHAB-tu1utUER_1Y1aR1f5E5UIMWIptdA-LZLHeE4DPqdgJeesDP75DF2z2Jclltjg: No such file or directory
(base) a-10-27-23-46:~ zoehaskellcraig\$ gcloud auth login --no-launch-browser
Go to the following link in your browser:

https://accounts.google.com/o/oauth2/auth?response_type=code&client_id=32555940559.apps.googleusercontent.com&redirect_uri=https%3A%2F%2Fsdk.cloud.google.com%2Fauthcode.html&scope=openid+https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fuserinfo.email+https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fcloud-platform+https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fappengine.admin+https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fsqlservice.login+https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fcompute+https%3A%2F%2Fwww.googleapis.com%2Fauth%2Faccounts.reauth&state=FX3Q0zB6XYz9sYYIRk8wmKjx35Xe24&prompt=consent&access_type=offline&code_challenge=ynt7Jr_2EFk80pE3TaC8oo0gS1j0pu0coahBRiPZ3B4&code_challenge_method=S256

Enter authorization code: 4/0AeaYSHBLhM3uafypdUI_-Jef0nAsgp-Yd7uRSxygq-mzuY6y6A0iKvABxpnc8S1Jdt8HBQ

You are now logged in as [zjh235@nyu.edu].
Your current project is [premier-data-bastion-1999]. You can change this setting by running:
\$ gcloud config set project PROJECT_ID
(base) a-10-27-23-46:~ zoehaskellcraig\$ eval `ssh-agent -s`
Agent pid 69557
(base) a-10-27-23-46:~ zoehaskellcraig\$
(base) a-10-27-23-46:~ zoehaskellcraig\$ ssh-add ~/.ssh/google_compute_engine
Identity added: /Users/zoehaskellcraig/.ssh/google_compute_engine (zoehaskellcraig@10-18-235-29.dynapool.wireless.nyu.edu)
(base) a-10-27-23-46:~ zoehaskellcraig\$
(base) a-10-27-23-46:~ zoehaskellcraig\$ gcloud compute ssh bastion-vm --ssh-flag="-A" --tunnel-through-iap --project=premier-data-bastion-1999
WARNING:

To increase the performance of the tunnel, consider installing NumPy. For instructions, please see https://cloud.google.com/iap/docs/using-tcp-forwarding#increasing_the_tcp_upload_bandwidth

This is a computer system managed by and property of New York University. This system may contain U.S. Government information, which is restricted to authorized users ONLY. Unauthorized access, use, misuse, or modification of this computer system or of the data contained herein or in transit to/from this system may subject the individual to administrative disciplinary actions, criminal and civil penalties. Users have no expectation of privacy. This system and equipment are subject to monitoring to ensure proper performance of applicable security features or procedures. Such monitoring may result in the acquisition, recording, and analysis of all data being communicated, transmitted, processed, or stored in this system by a user. If monitoring reveals possible evidence of criminal activity, such evidence may be provided to Law Enforcement Personnel.

ANYONE USING THIS SYSTEM EXPRESSLY CONSENTS TO SUCH MONITORING

Activate the web console with: systemctl enable --now cockpit.socket

Register this system with Red Hat Insights: insights-client --register
Create an account or view all your systems at <https://red.ht/insights-dashboard>
Last login: Mon Feb 26 18:47:55 2024 from 35.235.241.32
-bash-4.4\$
-bash-4.4\$ ssh 10.0.0.2

This is a computer system managed by and property of New York University. This system may contain U.S. Government information, which is restricted to authorized users ONLY. Unauthorized access, use, misuse, or modification of this computer system or of the data contained herein or in transit to/from this system may subject the individual to administrative disciplinary actions, criminal and civil penalties. Users have no expectation of privacy. This system and equipment are subject to monitoring to ensure proper performance of applicable security features or procedures. Such monitoring may result in the acquisition, recording, and analysis of all data being communicated, transmitted, processed, or stored in this system by a user. If monitoring reveals possible evidence of criminal activity, such evidence may be provided to Law Enforcement Personnel.

ANYONE USING THIS SYSTEM EXPRESSLY CONSENTS TO SUCH MONITORING

Activate the web console with: systemctl enable --now cockpit.socket

Register this system with Red Hat Insights: insights-client --register
Create an account or view all your systems at <https://red.ht/insights-dashboard>
Last login: Mon Feb 26 18:48:02 2024 from 11.0.0.2
[zjh235_nyu_edu@workspace-vm ~]\$ cd /scratch/
[zjh235_nyu_edu@workspace-vm scratch]\$ # Merging for respiratory distress dataset
[zjh235_nyu_edu@workspace-vm scratch]\$
[zjh235_nyu_edu@workspace-vm scratch]\$ setwd("/scratch/Premier/Raw_Data")
-bash: syntax error near unexpected token `"/scratch/Premier/Raw_Data"'
[zjh235_nyu_edu@workspace-vm scratch]\$
[zjh235_nyu_edu@workspace-vm scratch]\$ library(haven)
-bash: syntax error near unexpected token `haven'
[zjh235_nyu_edu@workspace-vm scratch]\$ library(dplyr)
-bash: syntax error near unexpected token `dplyr'
[zjh235_nyu_edu@workspace-vm scratch]\$ library(tidyr)
-bash: syntax error near unexpected token `tidyr'
[zjh235_nyu_edu@workspace-vm scratch]\$
[zjh235_nyu_edu@workspace-vm scratch]\$ # load file
[zjh235_nyu_edu@workspace-vm scratch]\$ print("loading files...")
-bash: syntax error near unexpected token `loading files...'
[zjh235_nyu_edu@workspace-vm scratch]\$ print("demo:")
-bash: syntax error near unexpected token `demo:'
[zjh235_nyu_edu@workspace-vm scratch]\$ t1 <- Sys.time()
-bash: syntax error near unexpected token `('

```

[zjh235_nyu_edu@workspace-vm scratch]$
[zjh235_nyu_edu@workspace-vm scratch]$
[zjh235_nyu_edu@workspace-vm scratch]$ t2 <- Sys.time()
-bash: syntax error near unexpected token `('
[zjh235_nyu_edu@workspace-vm scratch]$ print("Total time:", t2-t1)
-bash: syntax error near unexpected token `"Total time:",'
[zjh235_nyu_edu@workspace-vm scratch]$
[zjh235_nyu_edu@workspace-vm scratch]$
[zjh235_nyu_edu@workspace-vm scratch]$
[zjh235_nyu_edu@workspace-vm scratch]$ setwd("/scratch/Premier/Raw_Data/_paticd_diag")
-bash: syntax error near unexpected token `"/scratch/Premier/Raw_Data/_paticd_diag"'
[zjh235_nyu_edu@workspace-vm scratch]$ print("diagnosis:")
-bash: syntax error near unexpected token `"diagnosis:"'
[zjh235_nyu_edu@workspace-vm scratch]$ t1 <- Sys.time()
-bash: syntax error near unexpected token `('
[zjh235_nyu_edu@workspace-vm scratch]$
[zjh235_nyu_edu@workspace-vm scratch]$ files <- list.files()
-bash: syntax error near unexpected token `('
[zjh235_nyu_edu@workspace-vm scratch]$ dta_files <- files[grep("\\.dta$", files)]
-bash: syntax error near unexpected token `('
[zjh235_nyu_edu@workspace-vm scratch]$
[zjh235_nyu_edu@workspace-vm scratch]$ data <- NA
-bash: -: No such file or directory
[zjh235_nyu_edu@workspace-vm scratch]$
[zjh235_nyu_edu@workspace-vm scratch]$ for (f in dta_files){
-bash: syntax error near unexpected token `('
[zjh235_nyu_edu@workspace-vm scratch]$   if (stringr::str_detect(f, "wide-")){
-bash: syntax error near unexpected token `f,'
[zjh235_nyu_edu@workspace-vm scratch]$
[zjh235_nyu_edu@workspace-vm scratch]$     temp <- read.dta(f)
-bash: syntax error near unexpected token `('
[zjh235_nyu_edu@workspace-vm scratch]$
[zjh235_nyu_edu@workspace-vm scratch]$
[zjh235_nyu_edu@workspace-vm scratch]$       if (is.na(data)) { #first file
-bash: syntax error near unexpected token `data'
[zjh235_nyu_edu@workspace-vm scratch]$         data <- temp
-bash: -: No such file or directory
[zjh235_nyu_edu@workspace-vm scratch]$       } else {
-bash: syntax error near unexpected token `}'
[zjh235_nyu_edu@workspace-vm scratch]$         data <- rbind(data, temp)
-bash: syntax error near unexpected token `('
[zjh235_nyu_edu@workspace-vm scratch]$       }
-bash: syntax error near unexpected token `}'
[zjh235_nyu_edu@workspace-vm scratch]$
[zjh235_nyu_edu@workspace-vm scratch]$   }
-bash: syntax error near unexpected token `}'
[zjh235_nyu_edu@workspace-vm scratch]$ }
-bash: syntax error near unexpected token `}'
[zjh235_nyu_edu@workspace-vm scratch]$
[zjh235_nyu_edu@workspace-vm scratch]$ save(data, file = "nyu_allyears_diagnosis.RData")
-bash: syntax error near unexpected token `data,'
[zjh235_nyu_edu@workspace-vm scratch]$
[zjh235_nyu_edu@workspace-vm scratch]$
[zjh235_nyu_edu@workspace-vm scratch]$ t2 <- Sys.time()
-bash: syntax error near unexpected token `('
[zjh235_nyu_edu@workspace-vm scratch]$ print("Total time:", t2-t1)
-bash: syntax error near unexpected token `"Total time:",'
[zjh235_nyu_edu@workspace-vm scratch]$ R

```

R version 4.3.2 (2023-10-31) -- "Eye Holes"
 Copyright (C) 2023 The R Foundation for Statistical Computing
 Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
 You are welcome to redistribute it under certain conditions.
 Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
 Type 'contributors()' for more information and
 'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
 'help.start()' for an HTML browser interface to help.
 Type 'q()' to quit R.

```
> # Merging for respiratory distress dataset
```

```
setwd("/scratch/Premier/Raw_Data")
```

```
library(haven)
library(dplyr)
library(tidyr)
```

```
# load file
print("loading files...")
print("demo:")
t1 <- Sys.time()
```

```
t2 <- Sys.time()
print("Total time:", t2-t1)
```

```

setwd("/scratch/Premier/Raw_Data/_paticd_diag")
print("diagnosis:")
t1 <- Sys.time()

files <- list.files()
dta_files <- files[grep("\\.dta$", files)]

data <- NA

for (f in dta_files){
  if (stringr::str_detect(f, "wide_")){

    temp <- read.dta(f)

    if (is.na(data)) { #first file
      data <- temp
    } else {
      data <- rbind(data, temp)
    }

  }
}
print("Total time:", t2-t1)ears_diagnosis.RData")

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

  filter, lag

The following objects are masked from 'package:base':

  intersect, setdiff, setequal, union

[1] "loading files..."
[1] "demo:"
Error in print.default("Total time:", t2 - t1) :
  invalid printing digits 0
> library(haven)
library(dplyr)
library(tidyr)
> setwd("/scratch/Premier/Raw_Data/_paticd_diag")
print("diagnosis:")
t1 <- Sys.time()

files <- list.files()
dta_files <- files[grep("\\.dta$", files)]

data <- NA

for (f in dta_files){
  if (stringr::str_detect(f, "wide_")){

    temp <- read.dta(f)

    if (is.na(data)) { #first file
      data <- temp
    } else {
      data <- rbind(data, temp)
    }

  }
}

save(data, file = "nyu_allyears_diagnosis.RData")

t2 <- Sys.time()
print("Total time:", t2-t1)
[1] "diagnosis:"
Error in read.dta(f) : could not find function "read.dta"
> setwd("/scratch/Premier/Raw_Data/_paticd_diag")
print("diagnosis:")
t1 <- Sys.time()

files <- list.files()
dta_files <- files[grep("\\.dta$", files)]

data <- NA

for (f in dta_files){
  if (stringr::str_detect(f, "wide_")){

    temp <- read_dta(f)

    if (is.na(data)) { #first file
      data <- temp
    } else {
      data <- rbind(data, temp)
    }

  }
}

```

```

}

save(data, file = "nyu_allyears_diagnosis.RData")

t2 <- Sys.time()
print("Total time:", t2-t1)
[1] "diagnosis:"
Error in if (is.na(data)) { : the condition has length > 1
> setwd("/scratch/Premier/Raw_Data/_patcd_diag")
print("diagnosis:")
t1 <- Sys.time()

files <- list.files()
dta_files <- files[grepl("\\.dta$", files)]

init <- 1

for (f in dta_files){
  if (stringr::str_detect(f, "wide_")){
    print(f)
    temp <- read_dta(f)

    if (init == 1) { #first file
      data <- temp
      init = 0
    } else {
      data <- rbind(data, temp)
    }
  }
}

save(data, file = "nyu_allyears_diagnosis.RData")
[1] "diagnosis:"
[1] "wide_nyu_20171_patcd_diag.dta"
[1] "wide_nyu_20172_patcd_diag.dta"
[1] "wide_nyu_20173_patcd_diag.dta"
[1] "wide_nyu_20174_patcd_diag.dta"
[1] "wide_nyu_20181_patcd_diag.dta"
[1] "wide_nyu_20182_patcd_diag.dta"
[1] "wide_nyu_20183_patcd_diag.dta"
[1] "wide_nyu_20184_patcd_diag.dta"
[1] "wide_nyu_20191_patcd_diag.dta"
[1] "wide_nyu_20192_patcd_diag.dta"
[1] "wide_nyu_20193_patcd_diag.dta"
[1] "wide_nyu_20194_patcd_diag.dta"
[1] "wide_nyu_20201_patcd_diag.dta"
[1] "wide_nyu_20202_patcd_diag.dta"
[1] "wide_nyu_20203_patcd_diag.dta"
[1] "wide_nyu_20204_patcd_diag.dta"
[1] "wide_nyu_20211_patcd_diag.dta"
[1] "wide_nyu_20212_patcd_diag.dta"
[1] "wide_nyu_20213_patcd_diag.dta"
>
>
> q()
[zjh235_nyu_edu@workspace-vm scratch]$ R

R version 4.3.2 (2023-10-31) -- "Eye Holes"
Copyright (C) 2023 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> load(file = "nyu_allyears_diagnosis.RData")
Error in readChar(con, 5L, useBytes = TRUE) : cannot open the connection
In addition: Warning message:
In readChar(con, 5L, useBytes = TRUE) :
cannot open compressed file 'nyu_allyears_diagnosis.RData', probable reason 'No such file or directory'
> getwd()
[1] "/scratch"
> setwd("/scratch/Premier/Raw_Data/_patcd_diag")
> load(file = "nyu_allyears_diagnosis.RData")
>
>
>
> setwd("/scratch/Premier/Raw_Data/_patcpt")

library(dplyr)
library(tidyr)
library(haven)

```

```

print("cpt files..")

files <- list.files()
dta_files <- files[grepl("\\.dta$", files)]

init <- 1

for (f in dta_files){
  if (stringr::str_detect(f, "wide_")){
    print(f)
    temp <- read_dta(f)

    if (init == 1) { #first file
      all_cpt <- temp
      init = 0
    } else {
      all_cpt <- rbind(all_cpt, temp)
    }
  }
}

save(all_cpt, file = "nyu_allyears_cpt.RData")

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

  filter, lag

The following objects are masked from 'package:base':

  intersect, setdiff, setequal, union

```

```

[1] "cpt files.."
[1] "wide_nyu_20171_patcpt.dta"
[1] "wide_nyu_20172_patcpt.dta"
[1] "wide_nyu_20173_patcpt.dta"
[1] "wide_nyu_20174_patcpt.dta"
[1] "wide_nyu_20181_patcpt.dta"
[1] "wide_nyu_20182_patcpt.dta"
[1] "wide_nyu_20183_patcpt.dta"
[1] "wide_nyu_20184_patcpt.dta"
[1] "wide_nyu_20191_patcpt.dta"
[1] "wide_nyu_20192_patcpt.dta"
[1] "wide_nyu_20193_patcpt.dta"
[1] "wide_nyu_20194_patcpt.dta"
[1] "wide_nyu_20201_patcpt.dta"
[1] "wide_nyu_20202_patcpt.dta"
[1] "wide_nyu_20203_patcpt.dta"
[1] "wide_nyu_20204_patcpt.dta"
[1] "wide_nyu_20211_patcpt.dta"
[1] "wide_nyu_20212_patcpt.dta"
[1] "wide_nyu_20213_patcpt.dta"
>
>
>
> setwd("/scratch/Premier/Raw_Data/_paticd_proc")

library(dplyr)
library(tidyr)
library(haven)
print("proc files..")

```

```

files <- list.files()
dta_files <- files[grepl("\\.dta$", files)]

init <- 1

for (f in dta_files){
  if (stringr::str_detect(f, "wide_")){
    print(f)
    temp <- read_dta(f)

    if (init == 1) { #first file
      all_proc <- temp
      init = 0
    } else {
      all_proc <- rbind(all_proc, temp)
    }
  }
}

save(all_proc, file = "nyu_allyears_proc.RData")
[1] "proc files.."
[1] "wide_nyu_20171_paticd_proc.dta"
[1] "wide_nyu_20172_paticd_proc.dta"
[1] "wide_nyu_20173_paticd_proc.dta"
[1] "wide_nyu_20174_paticd_proc.dta"
[1] "wide_nyu_20181_paticd_proc.dta"

```

```

[1] "wide_nyu_20182_patidc_proc.dta"
[1] "wide_nyu_20183_patidc_proc.dta"
[1] "wide_nyu_20184_patidc_proc.dta"
[1] "wide_nyu_20191_patidc_proc.dta"
[1] "wide_nyu_20192_patidc_proc.dta"
[1] "wide_nyu_20193_patidc_proc.dta"
[1] "wide_nyu_20194_patidc_proc.dta"
[1] "wide_nyu_20201_patidc_proc.dta"
[1] "wide_nyu_20202_patidc_proc.dta"
[1] "wide_nyu_20203_patidc_proc.dta"
[1] "wide_nyu_20204_patidc_proc.dta"
[1] "wide_nyu_20211_patidc_proc.dta"
[1] "wide_nyu_20212_patidc_proc.dta"
[1] "wide_nyu_20213_patidc_proc.dta"
>
>
>
>
> q()
[zjh235_nyu_edu@workspace-vm scratch]$ cd /scratch/Premier/Raw_Data/
[zjh235_nyu_edu@workspace-vm Raw_Data]$ ls
Copyright.txt      nyu_aprdrd.txt    nyu_de_admin.txt  nyu_icdpoa.txt    nyu_patttype.txt  nyu_poorigin.txt  _patcpt           _patidc_proc
manifest_data_extract.txt  nyu_chgmstr.txt  nyu_disstat.txt   nyu_msdrmdc.txt  nyu_payor.txt     _patapdrdg        _patdemo          _patidc_proc.zip
nyu_admttype.txt    nyu_cptcode.txt  nyu_icdcode.txt   nyu_msdrd.txt    nyu_physpec.txt   _patbill          _patidc_diag      _readmit
[zjh235_nyu_edu@workspace-vm Raw_Data]$ cd _patdemo/
[zjh235_nyu_edu@workspace-vm _patdemo]$ ls
nyu_20171_patdemo.dta  nyu_20181_patdemo.txt  nyu_20192_patdemo.dta  nyu_20202_patdemo.txt  nyu_20213_patdemo.dta
nyu_20171_patdemo.txt  nyu_20182_patdemo.dta  nyu_20192_patdemo.txt  nyu_20203_patdemo.dta  nyu_20213_patdemo.txt
nyu_20172_patdemo.dta  nyu_20182_patdemo.txt  nyu_20193_patdemo.dta  nyu_20203_patdemo.txt  nyu_allyears_patdemo.dta
nyu_20172_patdemo.txt  nyu_20183_patdemo.dta  nyu_20193_patdemo.txt  nyu_20204_patdemo.dta  nyu_allyears_patdemo_INPATIENT_ONLY.dta
nyu_20173_patdemo.dta  nyu_20183_patdemo.txt  nyu_20194_patdemo.dta  nyu_20204_patdemo.txt  nyu_allyears_patdemo_OUTPATIENT_ONLY.dta
nyu_20173_patdemo.txt  nyu_20184_patdemo.dta  nyu_20194_patdemo.txt  nyu_20211_patdemo.dta
nyu_20174_patdemo.dta  nyu_20184_patdemo.txt  nyu_20201_patdemo.dta  nyu_20211_patdemo.txt
nyu_20174_patdemo.txt  nyu_20191_patdemo.dta  nyu_20201_patdemo.txt  nyu_20212_patdemo.dta
nyu_20181_patdemo.dta  nyu_20191_patdemo.txt  nyu_20202_patdemo.dta  nyu_20212_patdemo.txt
[zjh235_nyu_edu@workspace-vm _patdemo]$
[zjh235_nyu_edu@workspace-vm _patdemo]$
[zjh235_nyu_edu@workspace-vm _patdemo]$ R

R version 4.3.2 (2023-10-31) -- "Eye Holes"
Copyright (C) 2023 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> setwd("/scratch/Premier/Raw_Data")

library(haven)
library(dplyr)
library(tidyr)

print("loading .... demo")
print(Sys.time())
demo <- read_dta("/_patdemo/nyu_allyears_patdemo.dta")
print(Sys.time())

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union

[1] "loading .... demo"
[1] "2024-02-26 20:01:16 EST"
Error: '/_patdemo/nyu_allyears_patdemo.dta' does not exist.
> setwd("/scratch/Premier/Raw_Data")

library(haven)
library(dplyr)
library(tidyr)

print("loading .... demo")
print(Sys.time())
demo <- read_dta("_patdemo/nyu_allyears_patdemo.dta")
print(Sys.time())
[1] "loading .... demo"
[1] "2024-02-26 20:01:54 EST"
[1] "2024-02-26 20:06:41 EST"
>

```

```

>
>
>
>
>
>
>
>
>
> names(demo)
[1] "pat_key"           "medrec_key"       "disc_mon"         "disc_mon_seq"
[5] "adm_mon"          "prov_id"          "i_o_ind"          "pat_type"
[9] "ms_drg"           "ms_drg_mdc"       "point_of_origin"  "adm_type"
[13] "disc_status"      "mart_status"      "age"              "gender"
[17] "race"             "hispanic_ind"     "admphy_spec"      "adm_phy"
[21] "attphy_spec"      "att_phy"          "std_payor"        "los"
[25] "pat_charges"      "pat_cost"         "pat_fix_cost"     "pat_var_cost"
[29] "publish_type"
> print("loading .... diagnosis")
setwd("/scratch/Premier/Raw_Data/_paticd_diag")
print(Sys.time())
load(file = "nyu_allyears_diagnosis.RData")
print(Sys.time())

print("merging")
merged_data <- left_join(demo, data, by = "pat_key")
print(Sys.time())
[1] "loading .... diagnosis"
[1] "2024-02-26 20:10:46 EST"
[1] "2024-02-26 20:11:57 EST"
[1] "merging"
[1] "2024-02-26 20:12:40 EST"
> print("loading .... proc")
setwd("/scratch/Premier/Raw_Data/_paticd_proc")
print(Sys.time())
load(file = "nyu_allyears_proc.RData")
print(Sys.time())

print("merging")
merged_data <- left_join(merged_data, all_proc, by = "pat_key")
print(Sys.time())
[1] "loading .... proc"
[1] "2024-02-26 20:15:23 EST"
[1] "2024-02-26 20:15:29 EST"
[1] "merging"
[1] "2024-02-26 20:16:13 EST"
> colnames(merged_data)
[1] "pat_key"           "medrec_key"       "disc_mon"         "disc_mon_seq"
[5] "adm_mon"          "prov_id"          "i_o_ind"          "pat_type"
[9] "ms_drg"           "ms_drg_mdc"       "point_of_origin"  "adm_type"
[13] "disc_status"      "mart_status"      "age"              "gender"
[17] "race"             "hispanic_ind"     "admphy_spec"      "adm_phy"
[21] "attphy_spec"      "att_phy"          "std_payor"        "los"
[25] "pat_charges"      "pat_cost"         "pat_fix_cost"     "pat_var_cost"
[29] "publish_type"     "all_diagnoses"    "all_proc_codes"
> head(merged_data)
# A tibble: 6 × 31
  pat_key medrec_key disc_mon disc_mon_seq adm_mon prov_id i_o_ind pat_type
  <dbl>   <dbl>     <dbl>     <dbl>   <dbl>   <dbl> <chr>   <dbl>
1  26527999   31687923   2017102     1  2017102   2330  0      30
2  741707753  424745332   2017101     1  2017101    27  0      30
3  741709880  307441050   2017101     1  2016412   27  I      8
4  741710545  91939687   2017101     1  2017101   272  0      30
5  741710661  224810373   2017101     1  2017101   150  0      30
6  741713693  866425808   2017101     2  2017101   272  0      30
# 23 more variables: ms_drg <dbl>, ms_drg_mdc <dbl>, point_of_origin <chr>,
# adm_type <dbl>, disc_status <dbl>, mart_status <chr>, age <dbl>,
# gender <chr>, race <chr>, hispanic_ind <chr>, admphy_spec <dbl>,
# adm_phy <dbl>, attphy_spec <dbl>, att_phy <dbl>, std_payor <dbl>,
# los <dbl>, pat_charges <dbl>, pat_cost <dbl>, pat_fix_cost <dbl>,
# pat_var_cost <dbl>, publish_type <chr>, all_diagnoses <chr>,
# all_proc_codes <chr>
> head(merged_data$all_diagnoses)
[1] "N64.4"
[2] "C79.51, C79.51, C64.9, G89.3"
[3] "K81.0, R10.13, D68.61, E66.9, E87.6, G43.909, I10, K29.00, N39.0, Z68.26, Z86.718, Z89.512"
[4] "I42.9"
[5] "I48.91, Z00.00"
[6] "R06.00, D64.9, Z95.4"
>
>
>
> merged_data <- merged_data %>%
  mutate(ARDS = if_else(stringr::str_detect(all_diagnoses, "J80"), 1, 0))
> colnames(merged_data)
[1] "pat_key"           "medrec_key"       "disc_mon"         "disc_mon_seq"
[5] "adm_mon"          "prov_id"          "i_o_ind"          "pat_type"
[9] "ms_drg"           "ms_drg_mdc"       "point_of_origin"  "adm_type"
[13] "disc_status"      "mart_status"      "age"              "gender"
[17] "race"             "hispanic_ind"     "admphy_spec"      "adm_phy"
[21] "attphy_spec"      "att_phy"          "std_payor"        "los"
[25] "pat_charges"      "pat_cost"         "pat_fix_cost"     "pat_var_cost"
[29] "publish_type"     "all_diagnoses"    "all_proc_codes"   "ARDS"

```

```

> head(merged_data$ARDS)
[1] 0 0 0 0 0
> sum(merged_data$ARDS == 1)
[1] NA
> sum(merged_data$ARDS)
[1] NA
> sum(merged_data$ARDS, na.rm = TRUE)
[1] 99165
> length(merged_data$ARDS)
[1] 39482050
> merged_data <- merged_data %>%
  mutate(J96 = if_else(stringr::str_detect(all_diagnoses, "J96"), 1, 0))
> sum(merged_data$J96)
[1] NA
> sum(merged_data$J96, na.rm = TRUE)
[1] 4761493
>
>
> sum(is.na(merged_data$ARDS))
[1] 9
> sum(is.na(merged_data$J96))
[1] 9
> q()
[zjh235_nyu_edu@workspace-vm _patdemo]$ ls
nyu_20171_patdemo.dta  nyu_20181_patdemo.txt  nyu_20192_patdemo.dta  nyu_20202_patdemo.txt  nyu_20213_patdemo.dta
nyu_20171_patdemo.txt  nyu_20182_patdemo.dta  nyu_20192_patdemo.txt  nyu_20203_patdemo.dta  nyu_20213_patdemo.txt
nyu_20172_patdemo.dta  nyu_20182_patdemo.txt  nyu_20193_patdemo.dta  nyu_20203_patdemo.txt  nyu_allyears_patdemo.dta
nyu_20172_patdemo.txt  nyu_20183_patdemo.dta  nyu_20193_patdemo.txt  nyu_20204_patdemo.dta  nyu_allyears_patdemo_INPATIENT_ONLY.dta
nyu_20173_patdemo.dta  nyu_20183_patdemo.txt  nyu_20194_patdemo.dta  nyu_20204_patdemo.txt  nyu_allyears_patdemo_OUTPATIENT_ONLY.dta
nyu_20173_patdemo.txt  nyu_20184_patdemo.dta  nyu_20194_patdemo.txt  nyu_20211_patdemo.dta
nyu_20174_patdemo.dta  nyu_20184_patdemo.txt  nyu_20201_patdemo.dta  nyu_20211_patdemo.txt
nyu_20174_patdemo.txt  nyu_20191_patdemo.dta  nyu_20201_patdemo.txt  nyu_20212_patdemo.dta
nyu_20181_patdemo.dta  nyu_20191_patdemo.txt  nyu_20202_patdemo.dta  nyu_20212_patdemo.txt
[zjh235_nyu_edu@workspace-vm _patdemo]$ cd ..
[zjh235_nyu_edu@workspace-vm Raw_Data]$ ls
Copyright.txt  nyu_aprdrng.txt  nyu_de_admin.txt  nyu_icdpoa.txt  nyu_patttype.txt  nyu_poorigin.txt  _patcpt  _patid_proc
manifest_data_extract.txt  nyu_chgmstr.txt  nyu_disstat.txt  nyu_msdrmgdc.txt  nyu_payor.txt  _pataprdrng  _patdemo  _patid_proc.zip
nyu_admttype.txt  nyu_cptcode.txt  nyu_icdcode.txt  nyu_msdrng.txt  nyu_physpec.txt  _patbill  _patid_diag  _readmit
[zjh235_nyu_edu@workspace-vm Raw_Data]$ cd _pataprdrng/
[zjh235_nyu_edu@workspace-vm _pataprdrng]$ ls
nyu_20171_pataprdrng.dta  nyu_20174_pataprdrng.txt  nyu_20184_pataprdrng.dta  nyu_20193_pataprdrng.txt  nyu_20203_pataprdrng.dta  nyu_20212_pataprdrng.txt
nyu_20171_pataprdrng.txt  nyu_20181_pataprdrng.dta  nyu_20184_pataprdrng.txt  nyu_20194_pataprdrng.dta  nyu_20203_pataprdrng.txt  nyu_20213_pataprdrng.dta
nyu_20172_pataprdrng.dta  nyu_20181_pataprdrng.txt  nyu_20191_pataprdrng.dta  nyu_20194_pataprdrng.txt  nyu_20204_pataprdrng.dta  nyu_20213_pataprdrng.txt
nyu_20172_pataprdrng.txt  nyu_20182_pataprdrng.dta  nyu_20191_pataprdrng.txt  nyu_20201_pataprdrng.dta  nyu_20204_pataprdrng.txt  nyu_allyears_pataprdrng.dta
nyu_20173_pataprdrng.dta  nyu_20182_pataprdrng.txt  nyu_20192_pataprdrng.dta  nyu_20201_pataprdrng.txt  nyu_20211_pataprdrng.dta
nyu_20173_pataprdrng.txt  nyu_20183_pataprdrng.dta  nyu_20192_pataprdrng.txt  nyu_20202_pataprdrng.dta  nyu_20211_pataprdrng.txt
nyu_20174_pataprdrng.dta  nyu_20183_pataprdrng.txt  nyu_20193_pataprdrng.dta  nyu_20202_pataprdrng.txt  nyu_20212_pataprdrng.dta
[zjh235_nyu_edu@workspace-vm _pataprdrng]$ cd /scratch/Premier/Merged_Data/
[zjh235_nyu_edu@workspace-vm Merged_Data]$ cd /scratch/Premier/Raw_Data/
[zjh235_nyu_edu@workspace-vm Raw_Data]$ ls
Copyright.txt  nyu_aprdrng.txt  nyu_de_admin.txt  nyu_icdpoa.txt  nyu_patttype.txt  nyu_poorigin.txt  _patcpt  _patid_proc
manifest_data_extract.txt  nyu_chgmstr.txt  nyu_disstat.txt  nyu_msdrmgdc.txt  nyu_payor.txt  _pataprdrng  _patdemo  _patid_proc.zip
nyu_admttype.txt  nyu_cptcode.txt  nyu_icdcode.txt  nyu_msdrng.txt  nyu_physpec.txt  _patbill  _patid_diag  _readmit
[zjh235_nyu_edu@workspace-vm Raw_Data]$ cd /scratch/Premier/Merged_Data/
[zjh235_nyu_edu@workspace-vm Merged_Data]$ ls
nyu_allyears_demo+aprdrng_IP.dta  nyu_allyears_INPATIENT_demo+aprdrng+diag+cpt.dta  nyu_allyears_INPATIENT_demo+aprdrng+diag.dta
[zjh235_nyu_edu@workspace-vm Merged_Data]$ R

```

R version 4.3.2 (2023-10-31) -- "Eye Holes"
 Copyright (C) 2023 The R Foundation for Statistical Computing
 Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
 You are welcome to redistribute it under certain conditions.
 Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
 Type 'contributors()' for more information and
 'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
 'help.start()' for an HTML browser interface to help.
 Type 'q()' to quit R.

```

> #####
# PATAPR files

```

```

setwd("/scratch/Premier/Raw_Data/_pataprdrng")

```

```

library(dplyr)
library(tidyr)
library(haven)
print("demographic files..")

```

```

files <- list.files()
dta_files <- files[grep("\\.dta$", files)]

```

```

init <- 1

```

```

for (f in dta_files){
  if (!stringr::str_detect(f, "nyu_allyears")){

```



```

print(f)
temp <- read_dta(f)

if (init == 1) { #first file
  all_aprdrg <- temp
  init = 0
} else {
  all_aprdrg <- rbind(all_aprdrg, temp)
}
}
}

save(all_aprdrg, file = "nyu_allyears_aprdrg.RData")

#####
## Merge all files

setwd("/scratch/Premier/Raw_Data")

save(respiratory_demo_aprdrg_diag, file = "respiratory_demo_aprdrg_diag.RData")

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

  filter, lag

The following objects are masked from 'package:base':

  intersect, setdiff, setequal, union

[1] "demographic files.."
[1] "nyu_20171_pataprdrg.dta"
[1] "nyu_20172_pataprdrg.dta"
[1] "nyu_20173_pataprdrg.dta"
[1] "nyu_20174_pataprdrg.dta"
[1] "nyu_20181_pataprdrg.dta"
[1] "nyu_20182_pataprdrg.dta"
[1] "nyu_20183_pataprdrg.dta"
[1] "nyu_20184_pataprdrg.dta"
[1] "nyu_20191_pataprdrg.dta"
[1] "nyu_20192_pataprdrg.dta"
[1] "nyu_20193_pataprdrg.dta"
[1] "nyu_20194_pataprdrg.dta"
[1] "nyu_20201_pataprdrg.dta"
[1] "nyu_20202_pataprdrg.dta"
[1] "nyu_20203_pataprdrg.dta"
[1] "nyu_20204_pataprdrg.dta"
[1] "nyu_20211_pataprdrg.dta"
[1] "nyu_20212_pataprdrg.dta"
[1] "nyu_20213_pataprdrg.dta"
[1] "loading .... demo"
[1] "2024-02-26 20:42:00 EST"
[1] "2024-02-26 20:46:44 EST"
[1] "loading .... diagnosis"
[1] "2024-02-26 20:46:44 EST"
[1] "2024-02-26 20:47:52 EST"
[1] "merging"
[1] "2024-02-26 20:48:27 EST"
[1] "loading .... aprdrg"
[1] "2024-02-26 20:48:27 EST"
[1] "2024-02-26 20:48:44 EST"
[1] "merging"
[1] "2024-02-26 20:49:22 EST"
[1] "creating ARDS"
[1] "creating J96"
[1] "saving respiratory dataset"

```