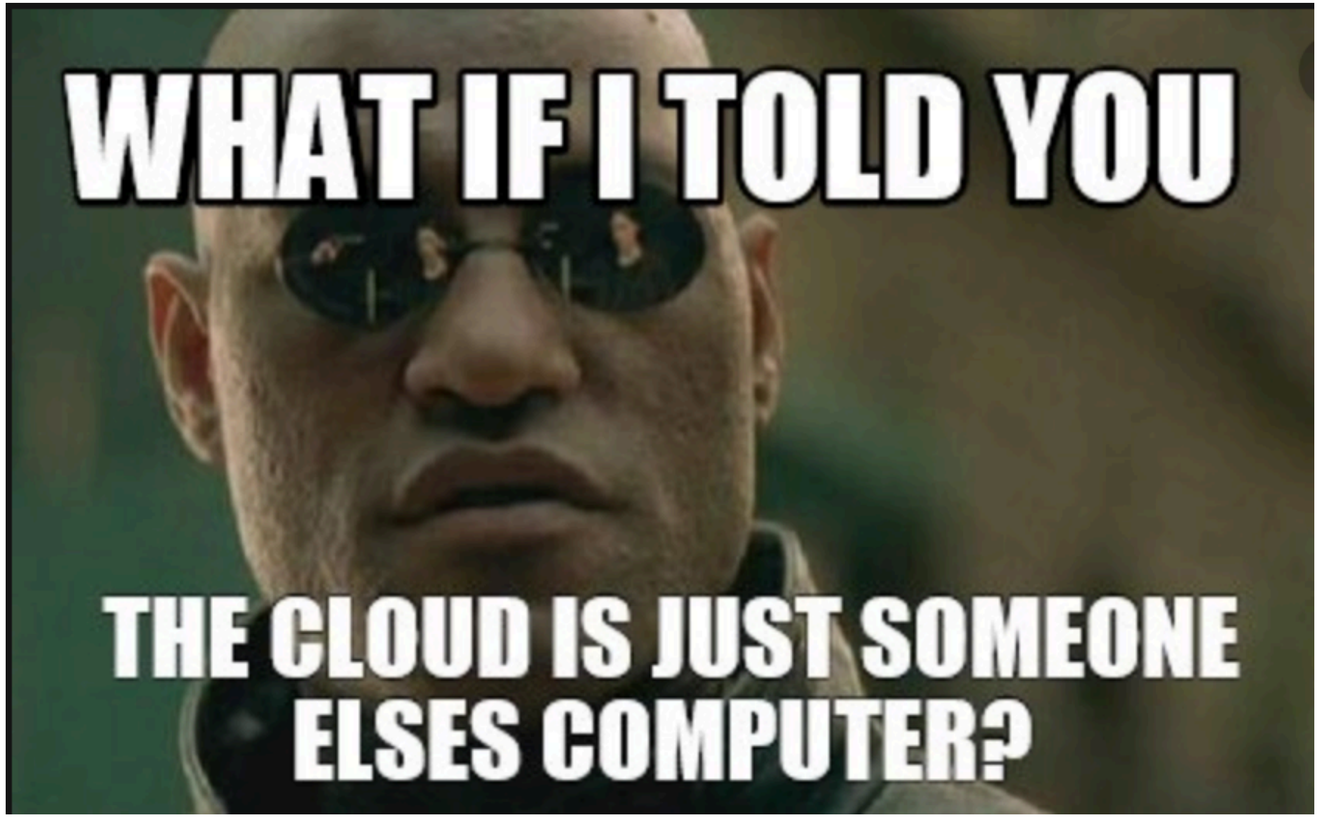


You don't have to download the Data:

Analysing ABCD on the cloud

Acknowledgement

Richard Beare

DevelopmentalImagingMCRI

Software from the Developmental Imaging Group, Murdoch Childrens Research Institute.

Melbourne, Australia.

Before you Start

1. ABCD lives in amazon web services.

   https://nda.nih.gov/training/module?trainingModuleId=training.cloud-access&slideId=slide.aws-creds

says that "To get started, email NDAHelp@mail.nih.gov and request that cloud access be added to your user account."

2. R Packages
   - *ROracle*
   - *dbplyr*
   - *tidyverse*

3. Create a miNDAR package on NDA website (and remember p/word)

```
## User name:             mls_mcri_147623
## Connect string:        (DESCRIPTION=(ADDRESS=(PROTOCOL=tcp)(HOST=mindarvpc.cqahbwk3l1mb.us-east-1.rd
## Server version:        12.1.0.2.0
## Server type:           Oracle RDBMS
## Results processed:     0
## OCI prefetch:          FALSE
## Bulk read:             1000
## Bulk write:            1000
## Statement cache size:  0
## Open results:          0
```

Environment

This works on a Mac (Mojave 10.14.6)
R version 3.5.3 (2019-03-11) -- "Great Truth”
RStudio Version 1.1.463

Ubuntu
How to query a oracle database.
https://thraxys.wordpress.com/2016/10/25/install-roracle-on-linux/

Function

```
## A helper function to set up a connection
  connectNDA <- function(username, password, host) {
    port <- 1521
    sid <- "ORCL"
    connectstring <- paste(
      "(DESCRIPTION=",
      "(ADDRESS=(PROTOCOL=tcp)(HOST=", host, ")(PORT=", port, "))",
      "(CONNECT_DATA=(SID=", sid, ")))", sep = "")
    driver <- dbDriver("Oracle")

    con <- dbConnect(driver, username=username, password=password, dbname=connectstring)
  return(con)
    }
```

```
 ABCD_pack <- connectNDA(username="mls_mcri_153226",
                         password="XXXXX",
                        host="mindarvpc.cqahbwk3l1mb.us-east-1.rds.amazonaws.com")
```

Check you are connected…

```
>   print(ABCD_pack)
User name:              mls_mcri_153226
Connect string:         (DESCRIPTION=(ADDRESS=(PROTOCOL=tcp)
(HOST=mindarvpc.cqahbwk3l1mb.us-east-1.rds.amazonaws.com)(PORT=1521))
(CONNECT_DATA=(SID=ORCL)))
Server version:         12.1.0.2.0
Server type:            Oracle RDBMS
Results processed:      0
OCI prefetch:           FALSE
Bulk read:              1000
Bulk write:             1000
Statement cache size:   0
Open results:           0
```

Warning - will drop you out if inactive

Start checking out data ~ dbListTable (DBI)

```
## ---- ListTables ----
    packagetables <- dbListTables(ABCD_pack)

 >     length(packagetables)                          # 288 packages
[1] 288

>     head(packagetables, 30)
 [1] "YALCS01"           "YACC01"            "TSTGODP401"        "TSTGODP301"
 [5] "TSTGODP201"        "TSTGODP101"        "TSTATROI01"        "STQ01"
 [9] "SSTTSTATP301"      "SSTTSTATP201"      "SSTTSTATP101"      "SSTSTATROI01"
[13] "SSTROI01"          "SRPF01"            "SPH01"             "S3_LINKS"
[17] "PSB01"             "PRQ01"             "PPS01"             "PMQ01"
[21] "PLUS01"            "PDTI401"           "PDTI01"            "PDEM02"
[25] "PASR01"            "PACKAGE_MESSAGE"   "PACC01"            "OMICS_EXPERIMENT"
[29] "NBACK_BWROI02"     "NBACK_BWROI01"
```

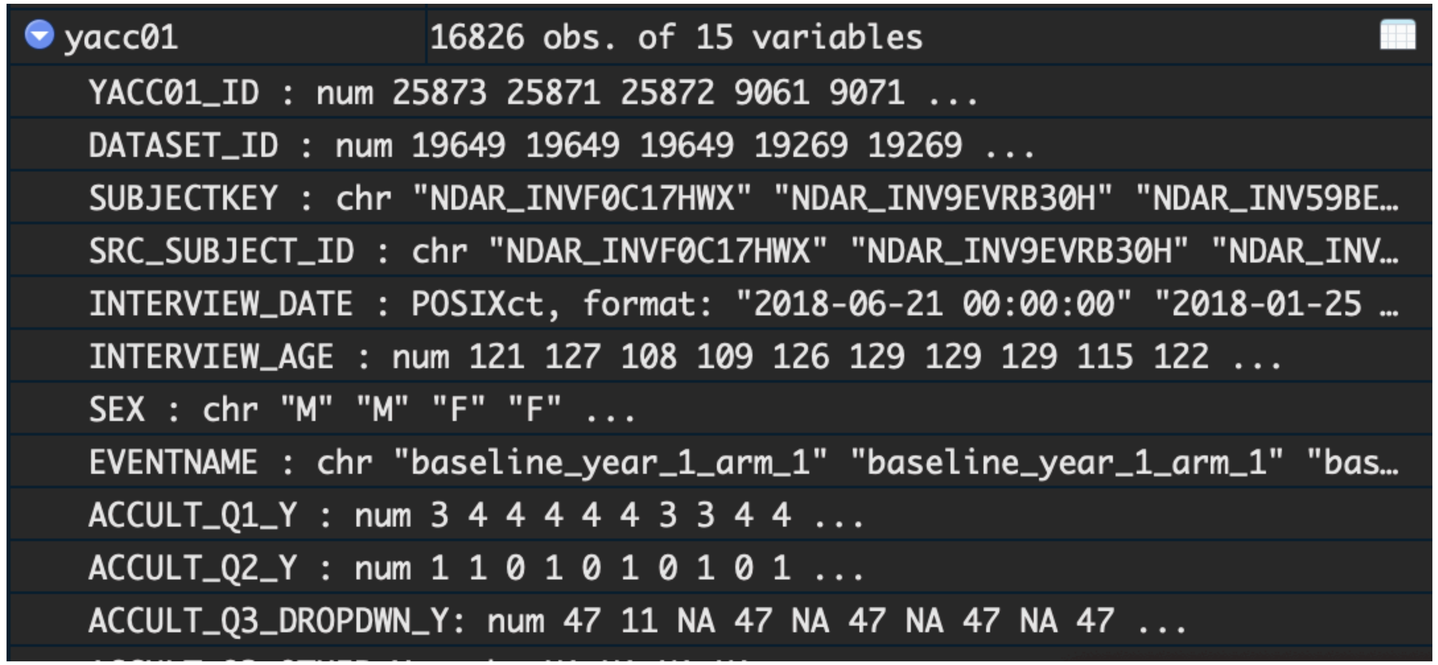
dbReadTable (R Oracle)

```
> pinfo <- dbReadTable(ABCD_pack, "PRQ01")

>     print(pinfo)
   PRQ01_ID DATASET_ID       SUBJECTKEY   SRC_SUBJECT_ID      INTERVIEW_DATE INTERVIEW_AGE
1      9048      19025 NDAR_INV0A6WVRZY NDAR_INV0A6WVRZY 2017-07-08 00:00:00           124
2      9053      19025 NDAR_INV0AEBMADL NDAR_INV0AEBMADL 2017-06-15 00:00:00           125
3      9064      19025 NDAR_INV0BAVEL0T NDAR_INV0BAVEL0T 2017-11-01 01:00:00           112
4      9077      19025 NDAR_INV0CBPKF6W NDAR_INV0CBPKF6W 2017-06-06 00:00:00           122
5      9080      19025 NDAR_INV0CCVJ39W NDAR_INV0CCVJ39W 2017-06-27 00:00:00           129
6      9082      19025 NDAR_INV0CFR33F0 NDAR_INV0CFR33F0 2018-05-11 00:00:00           108
7      9083      19025 NDAR_INV0CKA3YZX NDAR_INV0CKA3YZX 2017-05-25 00:00:00           124
8      9091      19025 NDAR_INV0D5J9T8P NDAR_INV0D5J9T8P 2017-05-18 00:00:00           128
9      9106      19025 NDAR_INV0E350J5D NDAR_INV0E350J5D 2018-01-02 01:00:00           123
```

dbReadTable (R Oracle) - Table Structure

```
> yacc01 <- dbReadTable(ABCD_pack, "YACC01")


> dim(yacc01)
[1] 16826    15
```

> object now in RStudio you can manipulate and save …



Tidyverse Stuff

acspsw03 [Socioeconomic]

```
> ## ---- Extract Relevant Data  ----

>     acspsw03 <- dbReadTable(ABCD_pack, "ACSPSW03")

>     colnames(acspsw03)
 [1] "ACSPSW03_ID"                  "DATASET_ID"                   "SUBJECTKEY"
 [4] "SRC_SUBJECT_ID"               "INTERVIEW_DATE"               "INTERVIEW_AGE"
 [7] "SEX"                          "EVENTNAME"                    "RACE_ETHNICITY"
[10] "REL_FAMILY_ID"                "REL_GROUP_ID"                 "REL_INGROUP_ORDER"
[13] "REL_RELATIONSHIP"             "REL_SAME_SEX"
"ACS_RAKED_PROPENSITY_SCORE"
[16] "STUDY_COHORT_NAME"
```

Tidyverse Stuff

acspsw03 [Socioeconomic]

```
> table(acspsw03$REL_GROUP_ID)

    1     2     3
11054   806    15


> family.structure <- group_by(acspsw03, REL_FAMILY_ID, REL_GROUP_ID)
> family.structure
# A tibble: 11,875 x 16
# Groups:   REL_FAMILY_ID, REL_GROUP_ID [10,802]
   ACSPSW03_ID DATASET_ID SUBJECTKEY SRC_SUBJECT_ID INTERVIEW_DATE      INTERVIEW_AGE SEX
         <dbl>      <dbl> <chr>      <chr>          <dttm>                     <dbl> <chr>
 1       50474      19649 NDAR_INVF… NDAR_INVF0C17… 2018-06-21 00:00:00           121 M
 2       50473      19649 NDAR_INV5… NDAR_INV59BE4… 2018-10-07 00:00:00           108 F
 3       50472      19649 NDAR_INV9… NDAR_INV9EVRB… 2018-01-25 01:00:00           127 M
 4       33649      19324 NDAR_INV0… NDAR_INV0A6WV… 2017-07-08 00:00:00           124 F
 5       33653      19324 NDAR_INV0… NDAR_INV0AAKG… 2018-03-18 01:00:00           131 F
 6       33657      19324 NDAR_INV0… NDAR_INV0APJM… 2017-06-22 00:00:00           126 F
 7       33661      19324 NDAR_INV0… NDAR_INV0AZFZ… 2018-03-06 01:00:00           116 F
 8       33662      19324 NDAR_INV0… NDAR_INV0B0HV… 2018-10-08 01:00:00           109 F
 9       33673      19324 NDAR_INV0… NDAR_INV0C1ED… 2017-08-14 00:00:00           125 F
```

Tidyverse Stuff

```
family.structure <- group_by(acspsw03, REL_FAMILY_ID, REL_GROUP_ID) %>%
                      summarise(gsize=n()) %>%
                      filter(gsize > 1)


# A tibble: 1,063 x 3
# Groups:   REL_FAMILY_ID [1,062]
   REL_FAMILY_ID REL_GROUP_ID gsize
           <dbl>        <dbl> <int>
 1            11            1     3
 2            32            1     2
 3            86            1     2
 4           111            1     2
 5           116            1     2
 6           156            1     2
 7           195            1     2
 8           246            1     2
 9           398            1     2
10           408            1     2
# … with 1,053 more rows
```

S3 links

```
> s3links <- dbReadTable(ABCD_pack, "S3_LINKS")

>   dim(s3links)
[1] 118009      1

>   head(s3links)
                                                                                                ENDPOINT
1 s3://NDAR_Central_1/submission_18712/NDAR_INVZ2GYNC07_abcd_fitbit_pilot_raw.tar.gz
2 s3://NDAR_Central_1/submission_18712/NDAR_INVDLXJMXAG_abcd_fitbit_pilot_raw.tar.gz
3 s3://NDAR_Central_1/submission_18712/NDAR_INV4UMAF27X_abcd_fitbit_pilot_raw.tar.gz
4 s3://NDAR_Central_1/submission_18712/NDAR_INV1MZ6AG9P_abcd_fitbit_pilot_raw.tar.gz
5 s3://NDAR_Central_1/submission_18712/NDAR_INV7ELLRXA6_abcd_fitbit_pilot_raw.tar.gz
6 s3://NDAR_Central_1/submission_18712/NDAR_INVHBGX2W56_abcd_fitbit_pilot_raw.tar.gz
```

S3 links

NOT MY CODE

```
 # Finding the MRI links
  s3links <- mutate(s3links,
                    ID=stringr::str_replace(basename(ENDPOINT), "([[:alnum:]]+)_.+", "\\1"),
                    ID2=stringr::str_replace(ID, "NDAR", "NDAR_"),
                    imagetype=stringr::str_remove(ENDPOINT, "s3:.+-MPROC-"),
                    imagetype=stringr::str_remove(imagetype, "_[[:digit:]]+\\.tgz"))


>      unique(s3links$imagetype)

 [152] "DTI"
 [153] "rsfMRI"
 [154] "nBack-fMRI"
 [155] "SST-fMRI"
 [156] "MID-fMRI"
 [157] "T1"
 [158] "T2"
```

twins_triplets_t1.txt

```
s3://NDAR_Central_2/submission_19173/NDARINVDBGLBFCJ_baselineYear1Arm1_ABCD-MPROC-T1_20180501125222.tgz
s3://NDAR_Central_2/submission_19173/NDARINV0WC5U4JA_baselineYear1Arm1_ABCD-MPROC-T1_20160920151851.tgz
s3://NDAR_Central_2/submission_19173/NDARINV4ZTRVWH5_baselineYear1Arm1_ABCD-MPROC-T1_20161226120533.tgz
s3://NDAR_Central_2/submission_19173/NDARINV6A16UBJN_baselineYear1Arm1_ABCD-MPROC-T1_20180415111943.tgz
s3://NDAR_Central_2/submission_19173/NDARINV58XVZFMC_baselineYear1Arm1_ABCD-MPROC-T1_20180626125511.tgz
s3://NDAR_Central_2/submission_19173/NDARINV4Z96NDW1_baselineYear1Arm1_ABCD-MPROC-T1_20170201181837.tgz
s3://NDAR_Central_2/submission_19173/NDARINV95L3764E_baselineYear1Arm1_ABCD-MPROC-T1_20180405161138.tgz
s3://NDAR_Central_2/submission_19173/NDARINV0RG4KBRB_baselineYear1Arm1_ABCD-MPROC-T1_20180416095409.tgz
s3://NDAR_Central_2/submission_19173/NDARINV851T3JYC_baselineYear1Arm1_ABCD-MPROC-T1_20180727124947.tgz
s3://NDAR_Central_2/submission_19173/NDARINV7AW78FJX_baselineYear1Arm1_ABCD-MPROC-T1_20170114101507.tgz
s3://NDAR_Central_2/submission_19173/NDARINV28X6J47M_baselineYear1Arm1_ABCD-MPROC-T1_20180307092627.tgz
s3://NDAR_Central_2/submission_19173/NDARINVBB64CLXY_baselineYear1Arm1_ABCD-MPROC-T1_20161222115941.tgz
s3://NDAR_Central_2/submission_19173/NDARINV831V8XME_baselineYear1Arm1_ABCD-MPROC-T1_20180406113503.tgz
s3://NDAR_Central_2/submission_19173/NDARINVDP1WJ2JC_baselineYear1Arm1_ABCD-MPROC-T1_20180609123804.tgz
s3://NDAR_Central_2/submission_21801/NDARINVECC0UT5E_baselineYear1Arm1_ABCD-MPROC-T1_20171021134438.tgz
s3://NDAR_Central_2/submission_21801/NDARINVB9WLDALN_baselineYear1Arm1_ABCD-MPROC-T1_20171006132108.tgz
s3://NDAR_Central_2/submission_21801/NDARINVXUKXXT12_baselineYear1Arm1_ABCD-MPROC-T1_20171003153428.tgz
s3://NDAR_Central_2/submission_21801/NDARINVM7LL5JGJ_baselineYear1Arm1_ABCD-MPROC-T1_20170803153034.tgz
s3://NDAR_Central_2/submission_21801/NDARINV3GFD7GXB_baselineYear1Arm1_ABCD-MPROC-T1_20161009112519.tgz
s3://NDAR_Central_2/submission_21801/NDARINVH8FRH0UK_baselineYear1Arm1_ABCD-MPROC-T1_20170910142131.tgz
s3://NDAR_Central_2/submission_21801/NDARINVX3GV1FW0_baselineYear1Arm1_ABCD-MPROC-T1_20170623085955.tgz
s3://NDAR_Central_2/submission_21801/NDARINVCT3X58XN_baselineYear1Arm1_ABCD-MPROC-T1_20161129115651.tgz
s3://NDAR_Central_2/submission_19161/NDARINV6XNN68K1_baselineYear1Arm1_ABCD-MPROC-T1_20180624131414.tgz
s3://NDAR_Central_2/submission_19161/NDARINV07BR8GMY_baselineYear1Arm1_ABCD-MPROC-T1_20180812123901.tgz
s3://NDAR_Central_2/submission_21801/NDARINVGPF6X87H_baselineYear1Arm1_ABCD-MPROC-T1_20161201164355.tgz
s3://NDAR_Central_2/submission_21801/NDARINV3JZ6Y396_baselineYear1Arm1_ABCD-MPROC-T1_20170930122503.tgz
s3://NDAR_Central_2/submission_21801/NDARINVPEM4YKYG_baselineYear1Arm1_ABCD-MPROC-T1_20170226103358.tgz
s3://NDAR_Central_2/submission_21801/NDARINVVKF7FYHA_baselineYear1Arm1_ABCD-MPROC-T1_20170909105322.tgz
s3://NDAR_Central_2/submission_21801/NDARINVYRW56GJ2_baselineYear1Arm1_ABCD-MPROC-T1_20161107135816.tgz
s3://NDAR_Central_2/submission_21801/NDARINV5LM4TFDU_baselineYear1Arm1_ABCD-MPROC-T1_20171122095721.tgz
s3://NDAR_Central_2/submission_21801/NDARINVUY1UGEA3_baselineYear1Arm1_ABCD-MPROC-T1_20161014124457.tgz
s3://NDAR_Central_2/submission_21801/NDARINVEDNJMEJ0_baselineYear1Arm1_ABCD-MPROC-T1_20161112122029.tgz
s3://NDAR_Central_2/submission_21801/NDARINVXN6HMGK8_baselineYear1Arm1_ABCD-MPROC-T1_20170906153621.tgz
s3://NDAR_Central_2/submission_19161/NDARINVHYDABU5J_baselineYear1Arm1_ABCD-MPROC-T1_20180623150356.tgz
s3://NDAR_Central_2/submission_19161/NDARINVDG5TKGEX_baselineYear1Arm1_ABCD-MPROC-T1_20171228123027.tgz
s3://NDAR_Central_2/submission_19161/NDARINVMBU1DVRD_baselineYear1Arm1_ABCD-MPROC-T1_20180127142219.tgz
s3://NDAR_Central_2/submission_19161/NDARINVL1UL6UM5_baselineYear1Arm1_ABCD-MPROC-T1_20180708135050.tgz
s3://NDAR_Central_2/submission_19161/NDARINVV3Z0YVUT_baselineYear1Arm1_ABCD-MPROC-T1_20180121142137.tgz
s3://NDAR_Central_2/submission_19161/NDARINVGMFL6ATJ_baselineYear1Arm1_ABCD-MPROC-T1_20180311132539.tgz
s3://NDAR_Central_2/submission_19161/NDARINV7M6BGL3E_baselineYear1Arm1_ABCD-MPROC-T1_20180731125145.tgz
s3://NDAR_Central_2/submission_19161/NDARINVD5P9PY9L_baselineYear1Arm1_ABCD-MPROC-T1_20180603133138.tgz
s3://NDAR_Central_2/submission_19161/NDARINVU6F8X730_baselineYear1Arm1_ABCD-MPROC-T1_20180809125148.tgz
```