Sisyphus Challenge Week 6: Strategus execution

Jenna/Jack

Sena/Chungsoo

Engineering open science systems that build trust into the real-world evidence generation and dissemination process

'System' required elements:
- Required phenotypes
- Analysis specifications
- Decision thresholds

Distributed data network, standardized to common data model

Network coordination

Data quality evaluation

Research question → Database diagnostics

Database diagnostics → Pass

Database diagnostics → Fail → STOP

Phenotype development and evaluation

Cohort definitions → Cohort diagnostics

Cohort diagnostics → Pass

Cohort diagnostics → Fail → STOP

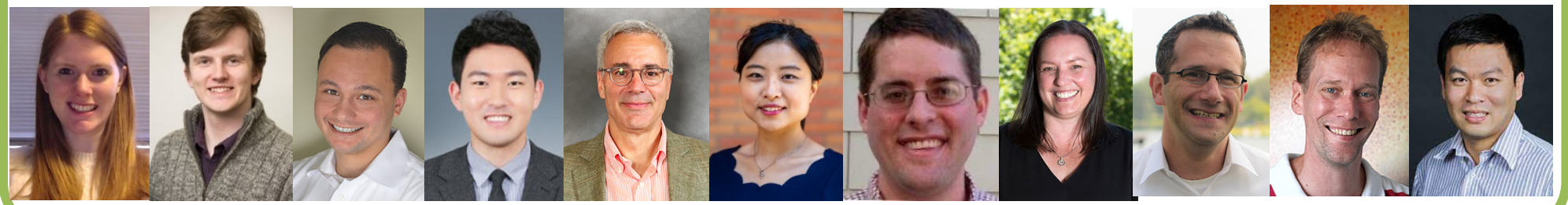
Analysis reliability evaluation

Analysis design choices → Study diagnostics

Study diagnostics → Pass

Study diagnostics → Fail → STOP

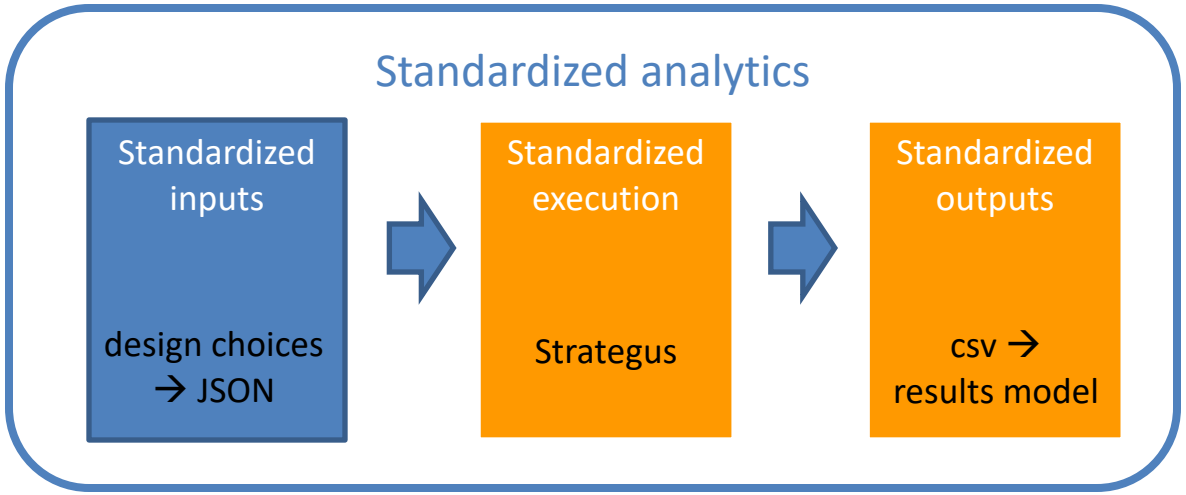
Week 6-8: Analysis execution, study diagnostics, evidence synthesis



Pass → Final unblinded results → Interface for exploration

Standardized analyses currently available within Strategus pipeline

- Characterization
  - Cohort diagnostics
  - Cohort features
  - Incidence rates
  - Time-to-event
  - Dechallenge / rechallenge
- Patient-level prediction
- Population-level effect estimation
  - Comparative cohort
  - Self-controlled case-series (SCCS)



OHDSI software for network studies

Strategus

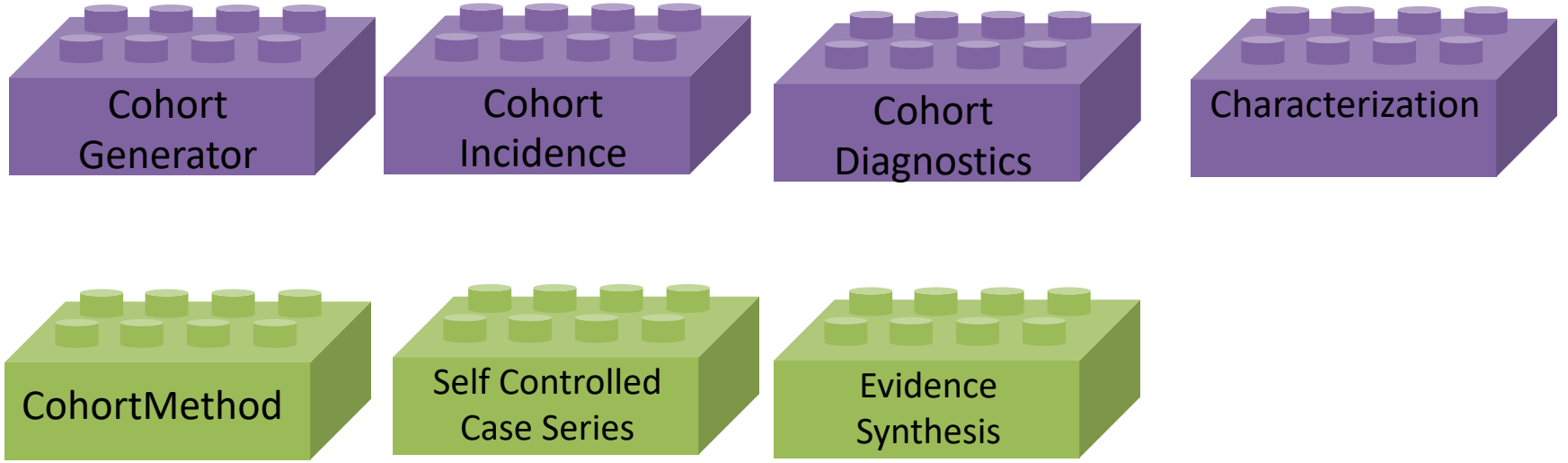
Pros

- Standardized representation of study inputs as JSON
- Modular composition of HADES analytics packages for study design
- Handles R environment configuration using renv

Reality

- Strategus is still in development

Strategus Modules



What is inside of a module? How is it different from a HADES package?

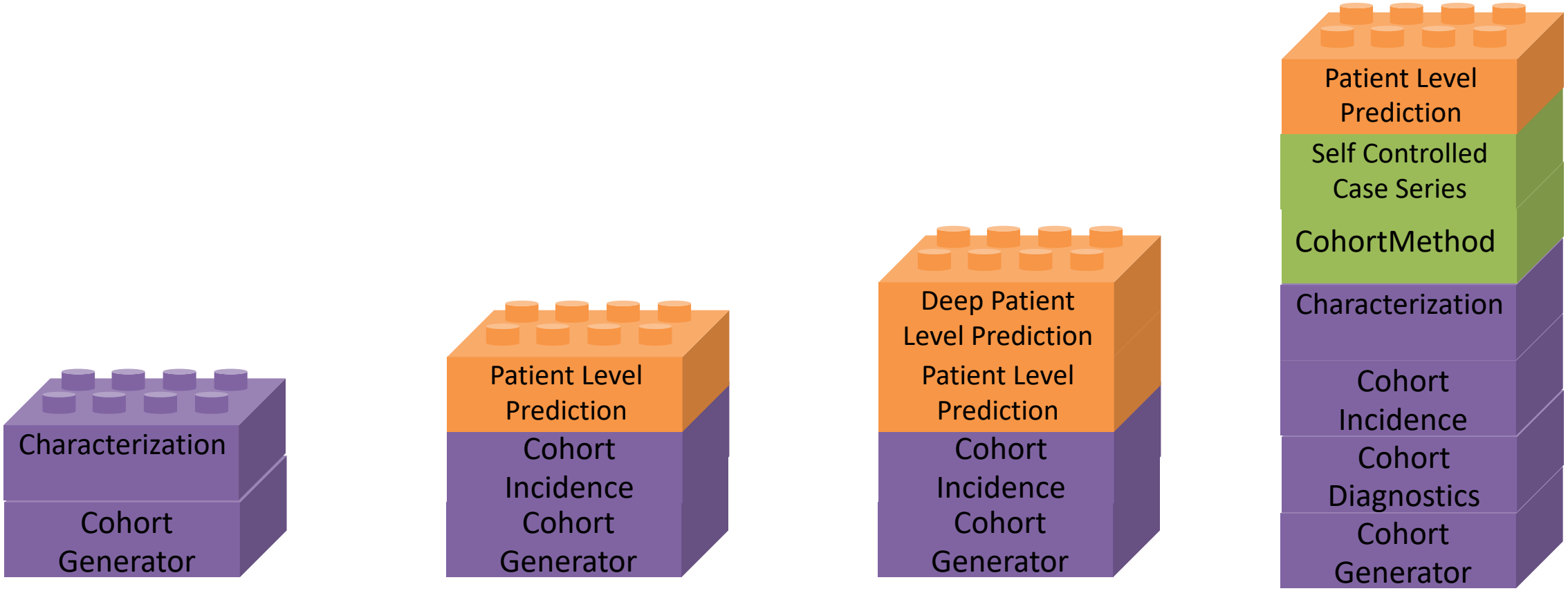
Strategus Modules

Strategus Module is an R Project that acts as a lightweight wrapper around one or more HADES packages. It has the following responsibilities:

- Declaring an renv.lock file for all R dependencies
- Providing functions for defining the module settings to store in the analysis specification JSON
- Implement an “execute” function to perform the analytical tasks based on module settings JSON using HADES packages
- Writing results as CSV files for sharing across the OHDSI network

Strategus Modules

- Strategus modules can be combined to accommodate various study designs.



Prerequisites to execute Strategus study

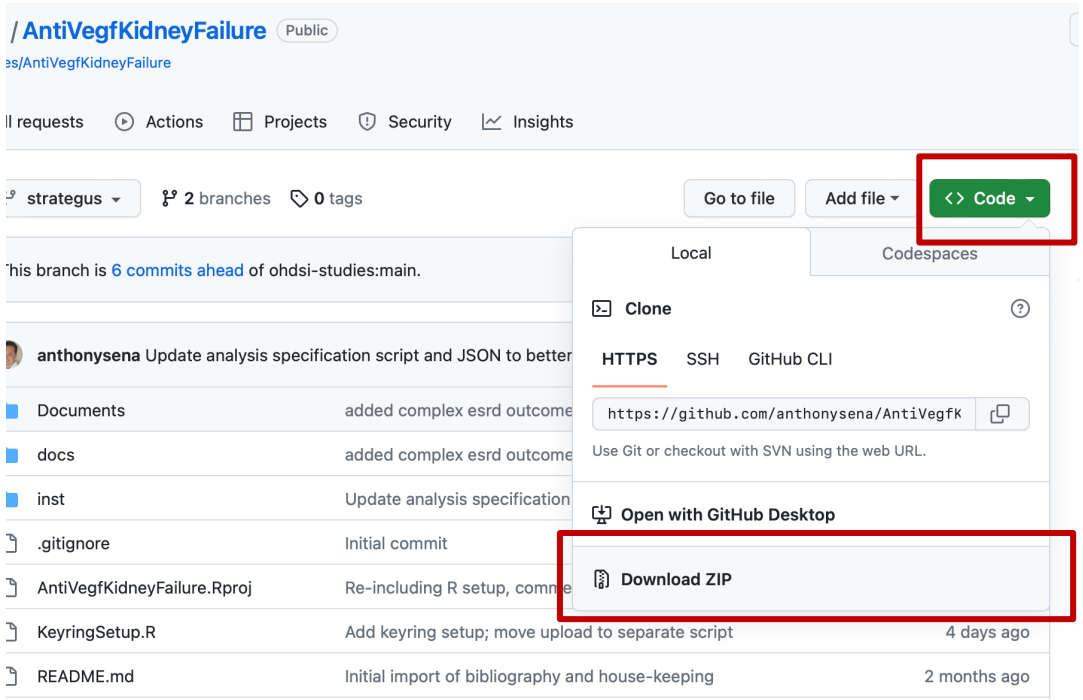
- **Database with OMOP CDM data**: convert your data to OMOP CDM and set up database server
- **Software:** Install R/R Studio + Java + setup GitHub personal access token (see HADES instructions: https://ohdsi.github.io/Hades/rSetup.html)

Strategus execution: 5 Steps in < 15 minutes

Download Project

Download project

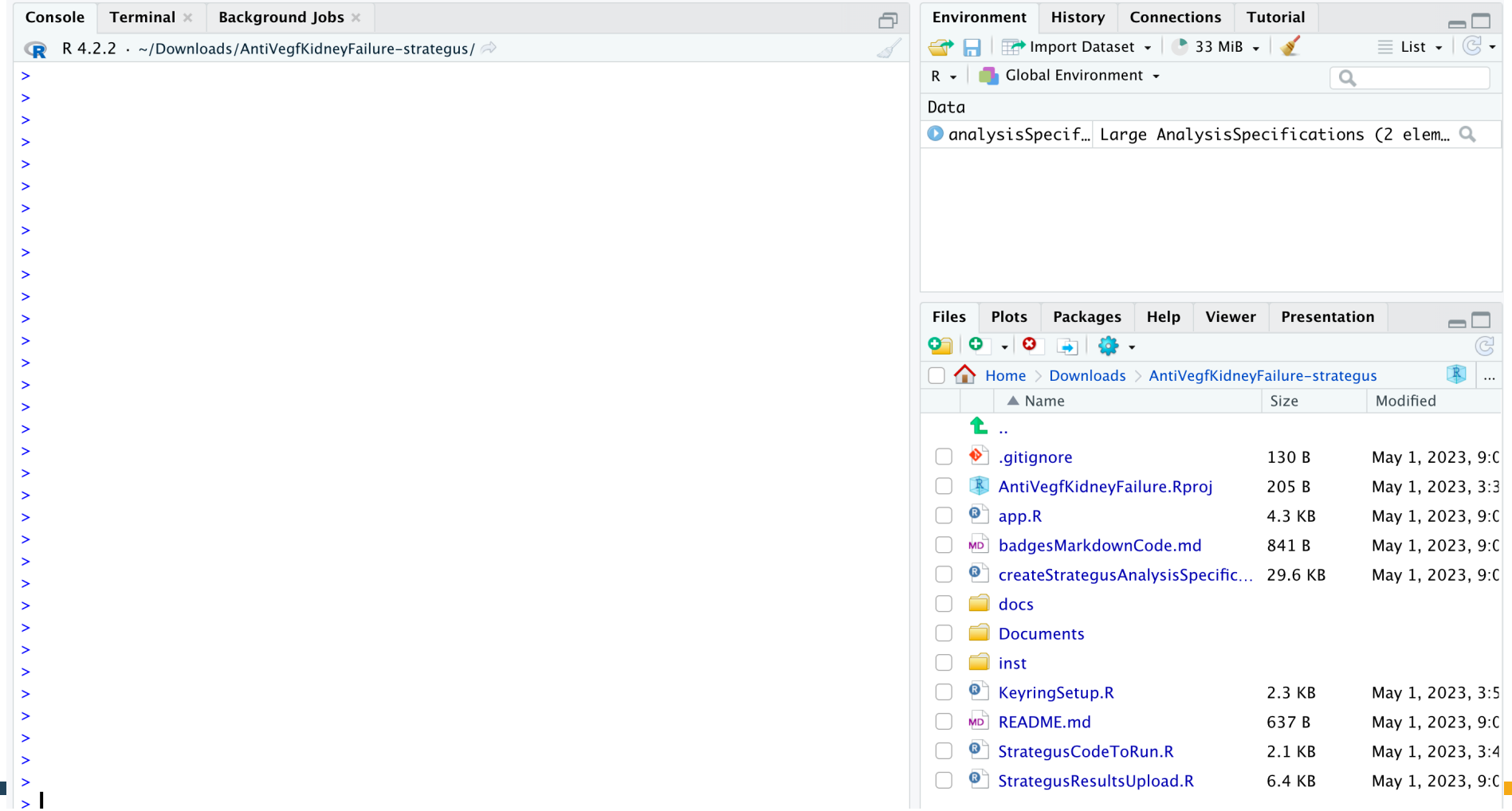
1. Go to study GitHub repo: https://github.com/ohdsi-studies/AntiVegfKidneyFailure
2. Download zipped folder
3. Unzip to local folder
4. Open folder and double click 'AntiVegfKidneyFailure.Rproj'



Configure

Configure local settings

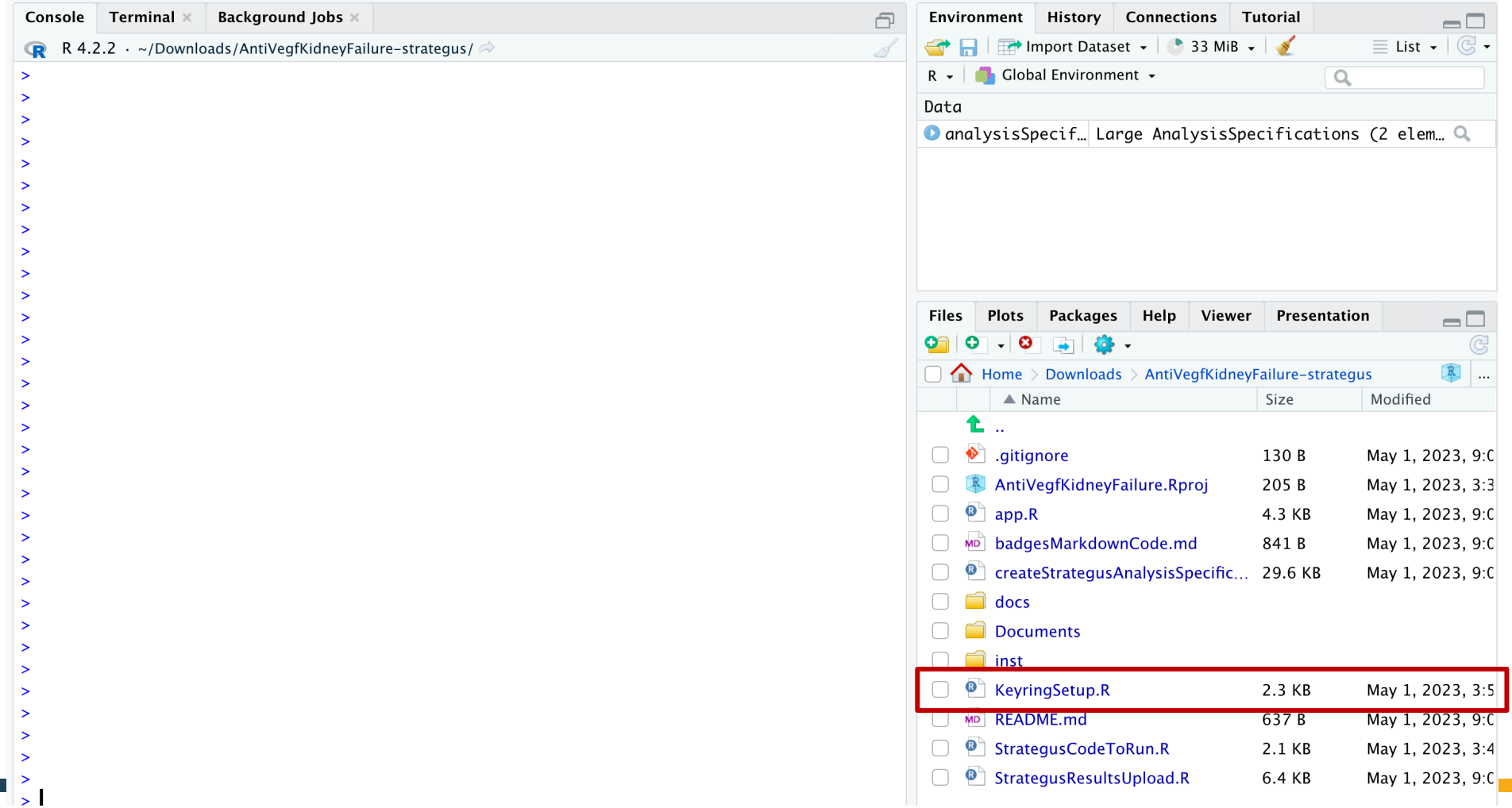
After opening AntiVegfKidneyFailure.Rproj you should see:



Configure

Configure local settings

Click on KeyringSetup.R to open the file in R Studio



Configure

Configure local settings

KeyringSetup.R

Source on Save | Run | Source

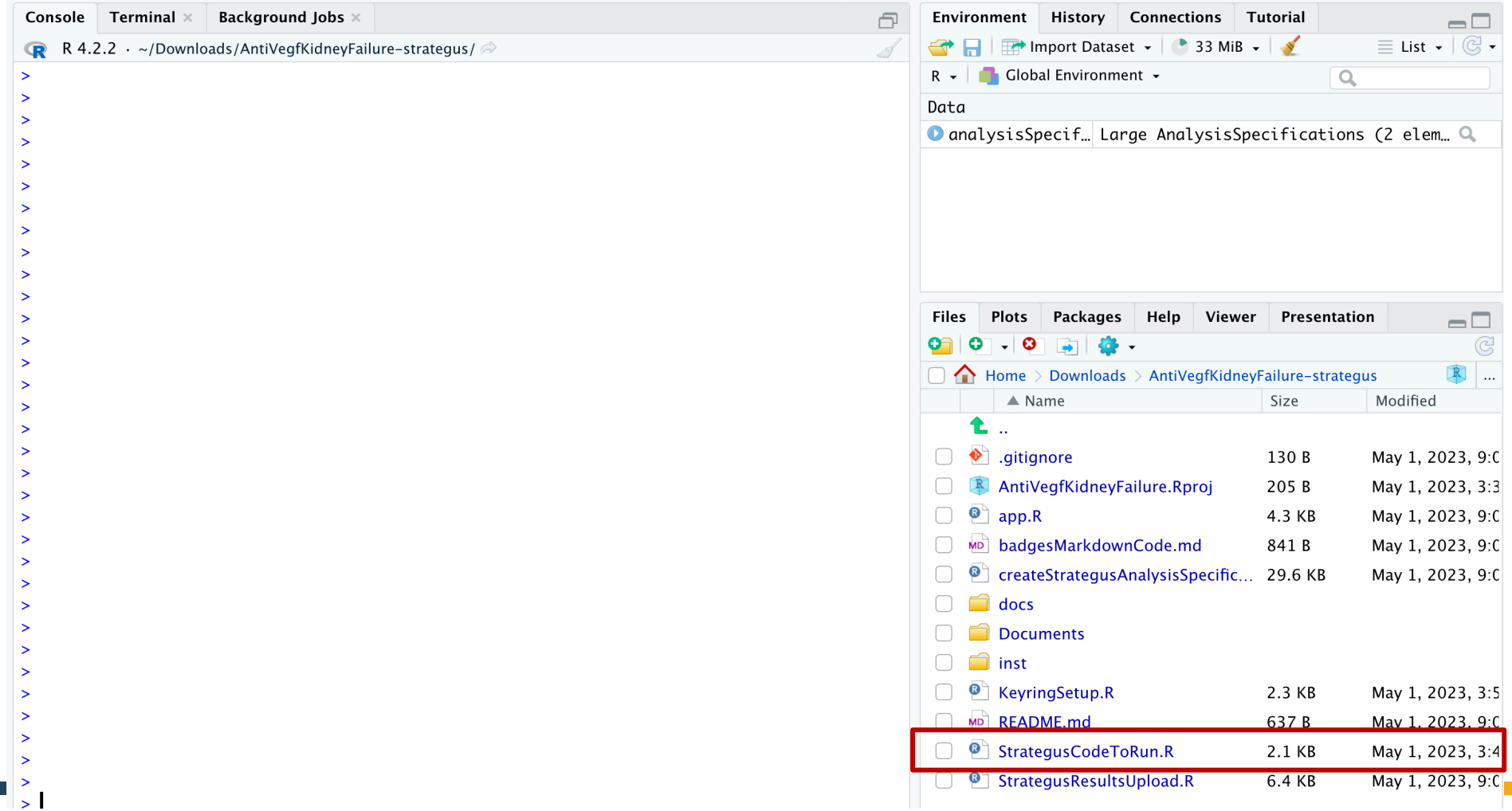
```r
# Install keyring - one time operation ---------
install.packages("keyring")

if (Sys.getenv("STRATEGUS_KEYRING_PASSWORD") == "") {
  # set keyring password by adding STRATEGUS_KEYRING_PASSWORD='sos' to renviron
  usethis::edit_r_environ()
  # then add STRATEGUS_KEYRING_PASSWORD='sos', save and close
  # Restart your R Session to confirm it worked
  stop("Please add STRATEGUS_KEYRING_PASSWORD='sos' to your .Renviron file
       via usethis::edit_r_environ() as instructed.")
}

# Provide your environment specific values ------
dbms <- "redshift"
connectionString <- "jdbc:redshift://your.server.goes.here:5439/your_cdm_database"
username <- "username-goes-here"
password = "password-goes-here"


# Run the rest to setup keyring ----------
```

1. Install keyring
2. Add strategus password to renviron
3. Configure database settings

Execute

Execute

- Open 'StrategusCodeToRun.R'



Execute

Execute

KeyringSetup.R | StrategusCodeToRun.R*

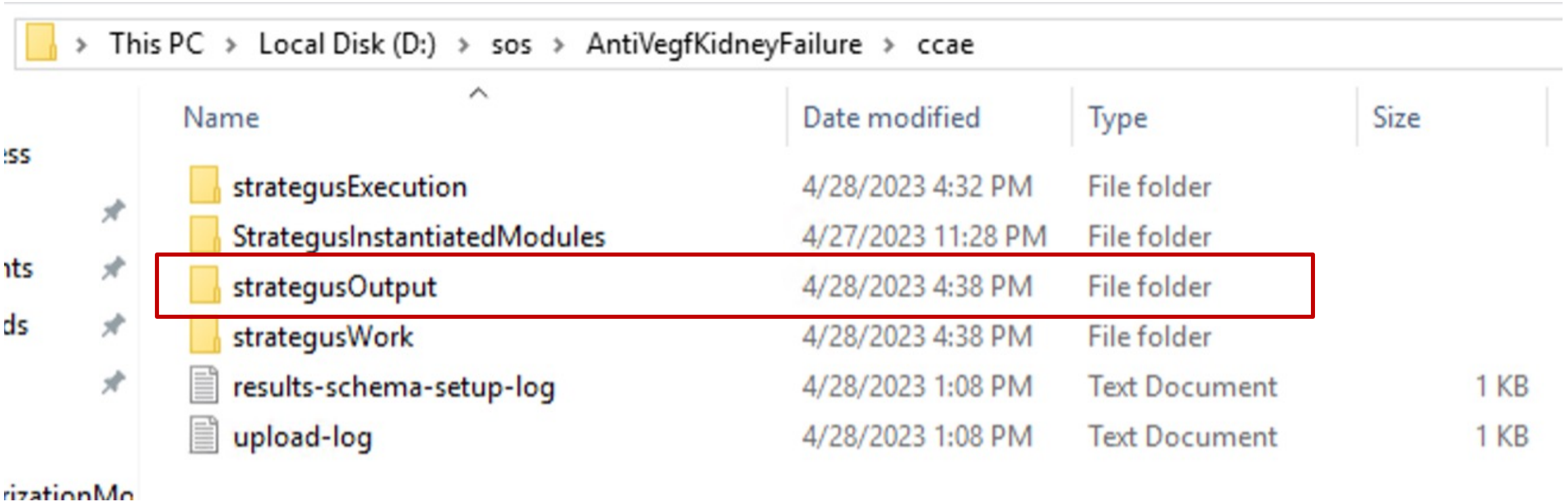
Source on Save | Run | Source

```r
# install the network package
install.packages('remotes')
remotes::install_github("OHDSI/Strategus", ref="results-upload")
library(Strategus)

##=========== START OF INPUTS ==========
connectionDetailsReference <- "Jmdc"
workDatabaseSchema <- 'scratch_asena5'
cdmDatabaseSchema <- 'cdm_jmdc_v2325'
outputLocation <- 'D:/git/anthonysena/AntiVegfKidneyFailure'
minCellCount <- 5
cohortTableName <- "sos_vegf_kf"

# the keyring entry should correspond to what you selected in KeyringSetup.R
connectionDetails = DatabaseConnector::createConnectionDetails(
  dbms = keyring::key_get("dbms", keyring = "sos-challenge"),
  connectionString = keyring::key_get("connectionString", keyring = "sos-challenge"),
  user = keyring::key_get("username", keyring = "sos-challenge"),
  password = keyring::key_get("password", keyring = "sos-challenge")
)

```

1. Install Strategus
2. Specify result output location and database settings
3. Specify connection details

Review CSV results

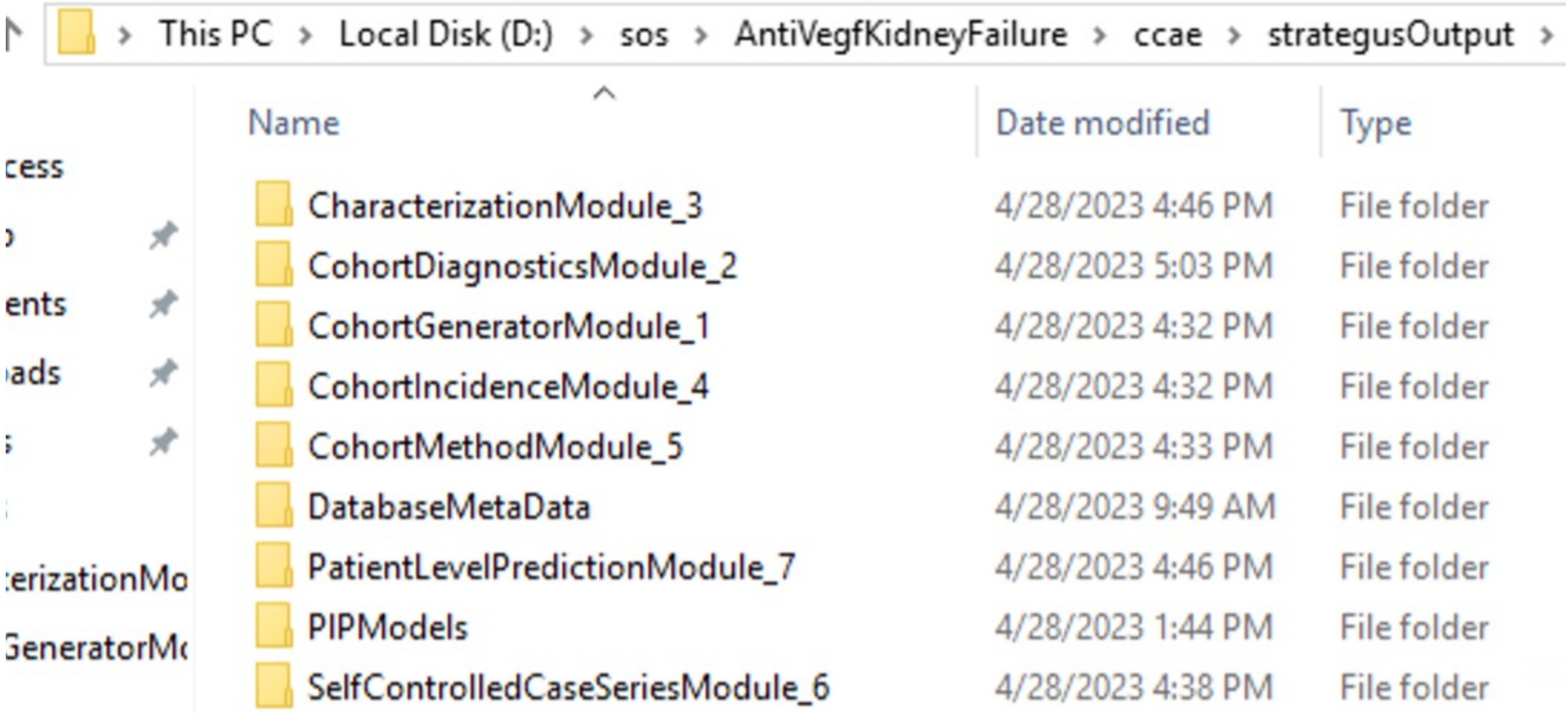
Review



The 'strategusOutput' folder contains the results for each module that was executed.

Inside strategusOutput folder

Review



Each module folder contains csv files

Review

CohortGenertorModule

This PC › Local Disk (D:) › sos › AntiVegfKidneyFailure › ccae › strategusOutput › CohortGeneratorModu

Name	Date modified	Type	Size
cg_cohort_count	4/28/2023 4:32 PM	CSV File	2 KB
cg_cohort_count_neg_ctrl	4/28/2023 4:32 PM	CSV File	2 KB
cg_cohort_definition	4/28/2023 4:32 PM	CSV File	271 KB
cg_cohort_generation	4/28/2023 9:58 AM	CSV File	4 KB
cg_cohort_inc_result	4/28/2023 4:32 PM	CSV File	1 KB
cg_cohort_inc_stats	4/28/2023 4:32 PM	CSV File	1 KB
cg_cohort_inclusion	4/28/2023 4:32 PM	CSV File	1 KB
cg_cohort_summary_stats	4/28/2023 4:32 PM	CSV File	1 KB
cohortGeneratorResults	4/28/2023 4:32 PM	Compressed (zipp...	31 KB
done	4/28/2023 4:32 PM	File	1 KB
log	4/28/2023 4:32 PM	Text Document	16 KB
resultsDataModelSpecification	4/28/2023 4:32 PM	CSV File	4 KB

CohortmethodModule

This PC › Local Disk (D:) › sos › AntiVegfKidneyFailure › ccae › strategusOutput › CohortMethodModule_5

Name	Date modified	Type	Size
cm_analysis	4/28/2023 4:32 PM	CSV File	7 KB
cm_attrition	4/28/2023 4:33 PM	CSV File	4 KB
cm_covariate	4/28/2023 4:33 PM	CSV File	6,468 KB
cm_covariate_analysis	4/28/2023 4:33 PM	CSV File	1 KB
cm_covariate_balance	4/28/2023 4:33 PM	CSV File	19 KB
cm_diagnostics_summary	4/28/2023 4:33 PM	CSV File	1 KB
cm_follow_up_dist	4/28/2023 4:33 PM	CSV File	1 KB
cm_interaction_result	4/28/2023 4:33 PM	CSV File	1 KB
cm_kaplan_meier_dist	4/28/2023 4:33 PM	CSV File	19 KB
cm_likelihood_profile	4/28/2023 4:33 PM	CSV File	350 KB
cm_preference_score_dist	4/28/2023 4:33 PM	CSV File	28 KB
cm_propensity_model	4/28/2023 4:33 PM	CSV File	63 KB
cm_result	4/28/2023 4:33 PM	CSV File	33 KB
cm_shared_covariate_balance	4/28/2023 4:33 PM	CSV File	5,914 KB
cm_target_comparator_outcome	4/28/2023 4:33 PM	CSV File	4 KB
done	4/28/2023 4:33 PM	File	1 KB
log	4/28/2023 4:33 PM	Text Document	176 KB
resultsDataModelSpecification	4/28/2023 4:33 PM	CSV File	19 KB

PatientLevelPredictionMo

This PC › Local Disk (D:) › sos › AntiVegfKidneyFailure › ccae › strategusOutput › PatientLevelPredictionModule_7

Name	Date modified	Type	Size
models	4/28/2023 12:22 PM	File folder	
attrition	4/28/2023 4:46 PM	CSV File	1 KB
calibration_summary	4/28/2023 4:46 PM	CSV File	147 KB
cohort_definition	4/28/2023 4:46 PM	CSV File	9 KB
cohorts	4/28/2023 4:46 PM	CSV File	1 KB
covariate_settings	4/28/2023 4:46 PM	CSV File	1 KB
covariate_summary	4/28/2023 4:46 PM	CSV File	27 KB
database_details	4/28/2023 4:46 PM	CSV File	1 KB
database_meta_data	4/28/2023 4:46 PM	CSV File	1 KB
demographic_summary	4/28/2023 4:46 PM	CSV File	36 KB
diagnostic_designs	4/28/2023 4:46 PM	CSV File	1 KB
diagnostic_outcomes	4/28/2023 4:46 PM	CSV File	18 KB
diagnostic_participants	4/28/2023 4:46 PM	CSV File	3 KB
diagnostic_predictors	4/28/2023 4:46 PM	CSV File	20 KB
diagnostic_summary	4/28/2023 4:46 PM	CSV File	1 KB
diagnostics	4/28/2023 4:46 PM	CSV File	1 KB
done	4/28/2023 4:46 PM	File	1 KB
evaluation_statistics	4/28/2023 4:46 PM	CSV File	8 KB
feature_engineering_settings	4/28/2023 4:46 PM	CSV File	1 KB
log	4/28/2023 4:46 PM	Text Document	186 KB
model_designs	4/28/2023 4:46 PM	CSV File	1 KB
model_settings	4/28/2023 4:46 PM	CSV File	3 KB
models	4/28/2023 4:46 PM	CSV File	1 KB
performances	4/28/2023 4:46 PM	CSV File	1 KB
plp_data_settings	4/28/2023 4:46 PM	CSV File	1 KB
population_settings	4/28/2023 4:46 PM	CSV File	1 KB
prediction_distribution	4/28/2023 4:46 PM	CSV File	4 KB
recalibrations	4/28/2023 4:46 PM	CSV File	1 KB
sample_settings	4/28/2023 4:46 PM	CSV File	1 KB
split_settings	4/28/2023 4:46 PM	CSV File	1 KB
tars	4/28/2023 4:46 PM	CSV File	1 KB
threshold_summary	4/28/2023 4:46 PM	CSV File	498 KB
tidy_covariates_settings	4/28/2023 4:46 PM	CSV File	1 KB

Share results

Share

main AntiVegfKidneyFailure / ShareResults.R

anthonysena Fix FTP folder name

1 contributor

57 lines (53 sloc) | 2.52 KB

```r
##============ START OF INPUTS ==========
connectionDetailsReference <- "Jmdc"
outputLocation <- 'D:/git/anthonysena/AntiVegfKidneyFailure'
# For uploading the results. You should have received the key file from the study coordinator:
keyFileName <- "[location where you are storing: e.g. ~/keys/study-data-site-covid19.dat]"
userName <- "study-data-site-covid19"
##============ END OF INPUTS ==========
#################################
# DO NOT MODIFY BELOW THIS POINT
#################################

rootFTPFolder <- function() {
  return("/sos-anti-vegf-kf/")
}

zipResults <- function(outputLocation, connectionDetailsReference) {
  resultsFolder <- file.path(outputLocation, connectionDetailsReference, "strategusOutput")
  zipFileName <- file.path(resultsFolder, paste0("Results_", connectionDetailsReference, ".zip"))
  DatabaseConnector::createZipFile(
```

To share csv results:

Add database reference
Add location of results

Get keyFileName and userName from study coordinator

5 Step Summary

- **Download:** Download zip in github page (https://github.com/ohdsi-studies/AntiVegfKidneyFailure), unzip folder in local directory and open R project (AntiVegfKidneyFailure.Rproj).
- **Configure:** Open KeyringSetup.R to configure database connection settings and keyring password using keyring
- **Execute:** Open StrategusCodeToRun.R to edit local output directory and database parameters and then run study using those settings
- **Review:** explore csv files in StrategusOutput and/or launch shiny
- **Share:** open ShareResults.R to export result files to OHDSI

Viewing shiny app

- First you need to upload the results to a database, we recommend sqlite for local viewing.

- To upload results into local sqlite database open the file "StrategusResultsUpload.R"

You only need to specify two inputs then run the script to insert the results into a local sqlite database

strategus ▾ AntiVegfKidneyFailure / StrategusResultsUpload.R

anthonysena Updates for upload and shiny

1 contributor

163 lines (151 sloc) | 6.4 KB

```
1   ##=========== START OF INPUTS ==========
2
3   connectionDetailsReference <- "Jmdc"
4   outputLocation <- 'D:/git/anthonysena/AntiVegfKidneyFailure'
5
6   ##=========== END OF INPUTS ==========
7   ######################################
```

Set the connectionDetailsReference to a reference for your OMOP CDM database

Set the outputLocation to a folder location on your hard drive where you can read/write from

Your data are now ready for the shiny app

- Launch shiny app open the file "app.R"

strategus / AntiVegfKidneyFailure / app.R

anthonysena Updates for upload and shiny

1 contributor

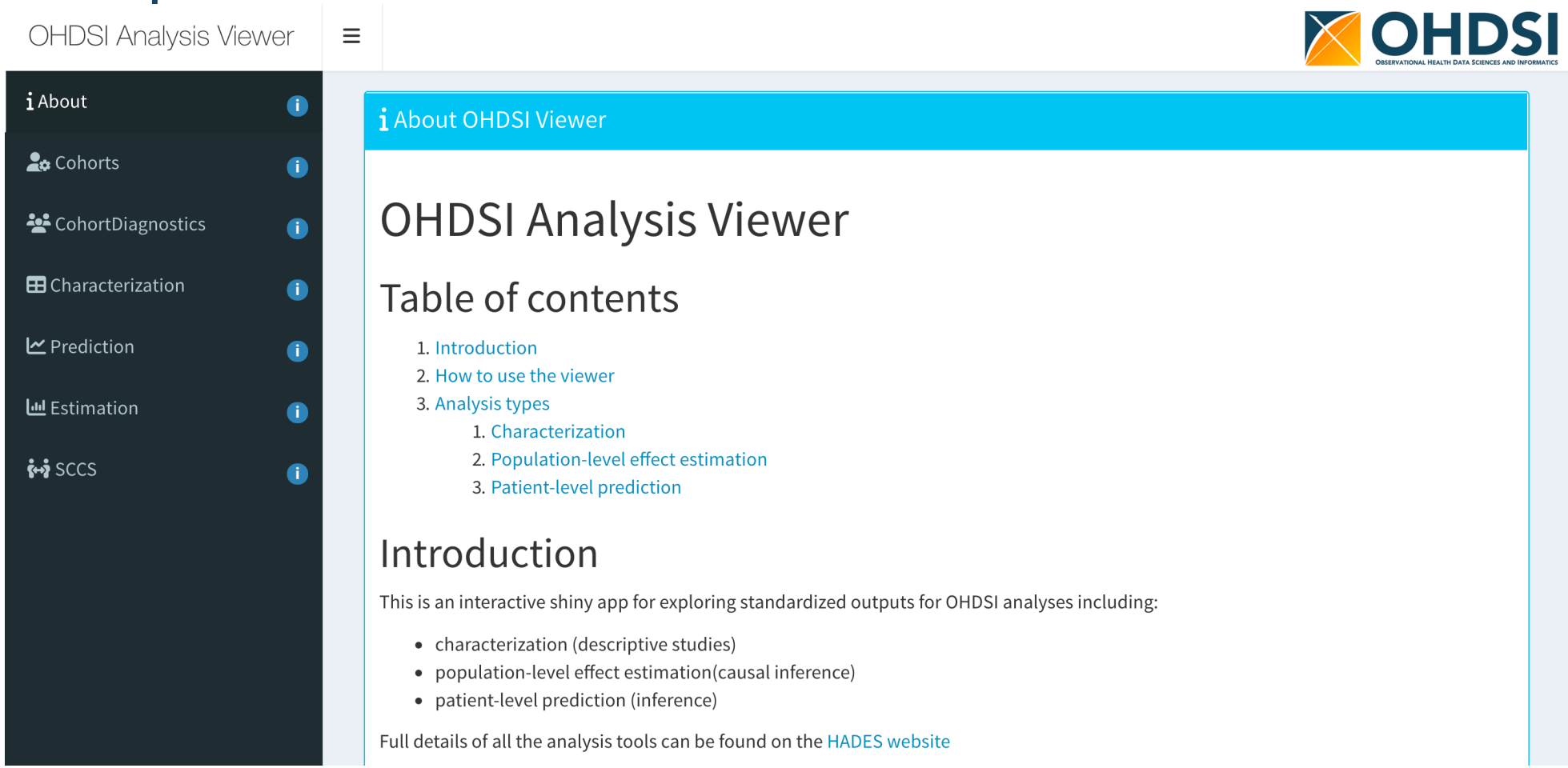
132 lines (119 sloc) | 4.28 KB

```
1   # View Strategus results in the results database
2   # remotes::install_github("ohdsi/ShinyAppBuilder", ref = "develop")
3   # remotes::install_github("ohdsi/OhdsiShinyModules", ref = "develop")
4
5   library(dplyr)
6   library(ShinyAppBuilder)
7   library(markdown)
8
9   ##============ START OF INPUTS ==========
10
11  connectionDetailsReference <- "Jmdc"
12  outputLocation <- 'D:/git/anthonysena/AntiVegfKidneyFailure'
13
14  ##============ END OF INPUTS ==========
```

Only two inputs are required (the same values used when creating the local sqlite database on slide 22)

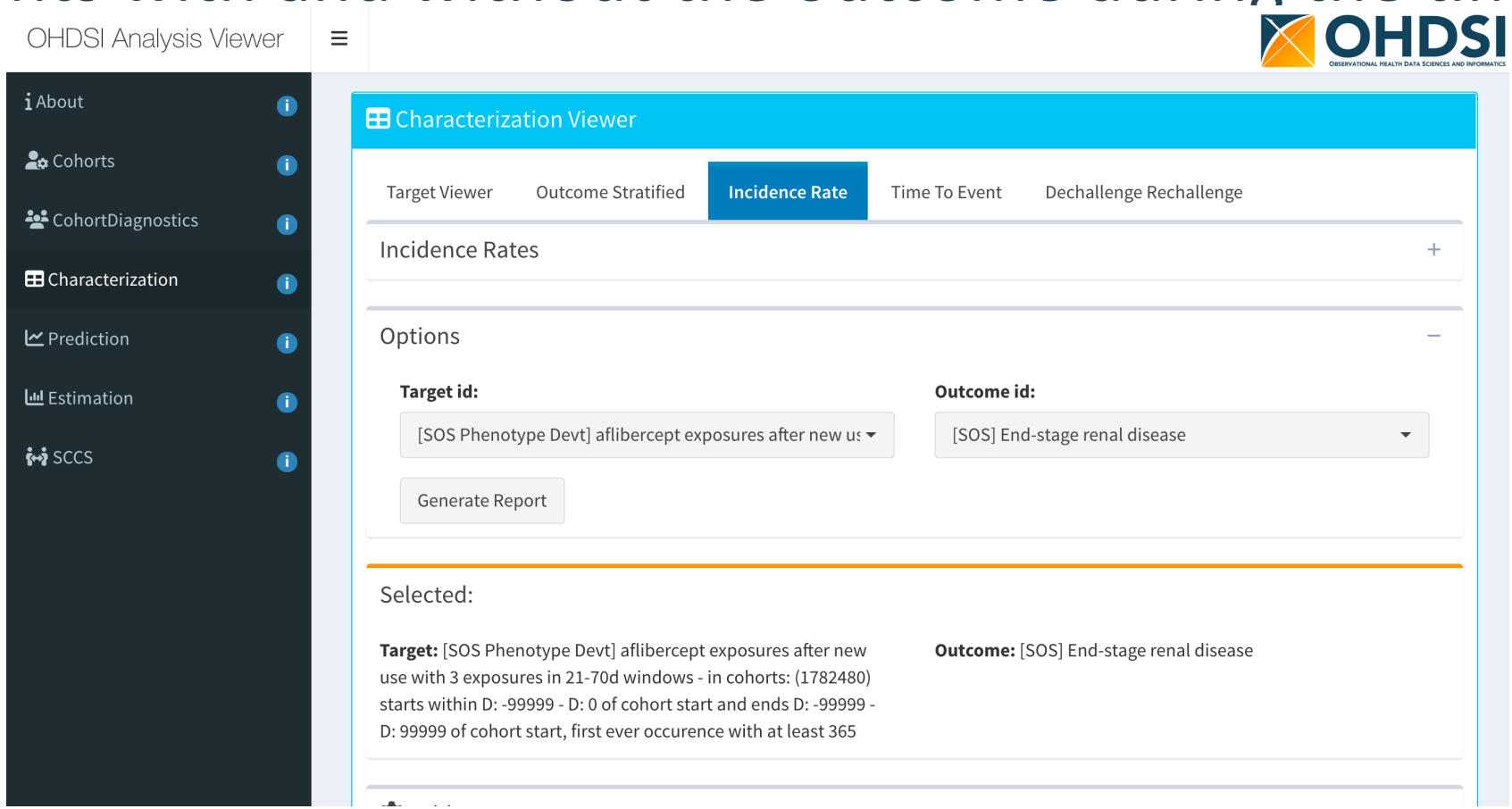
Shiny App

Let's explore...



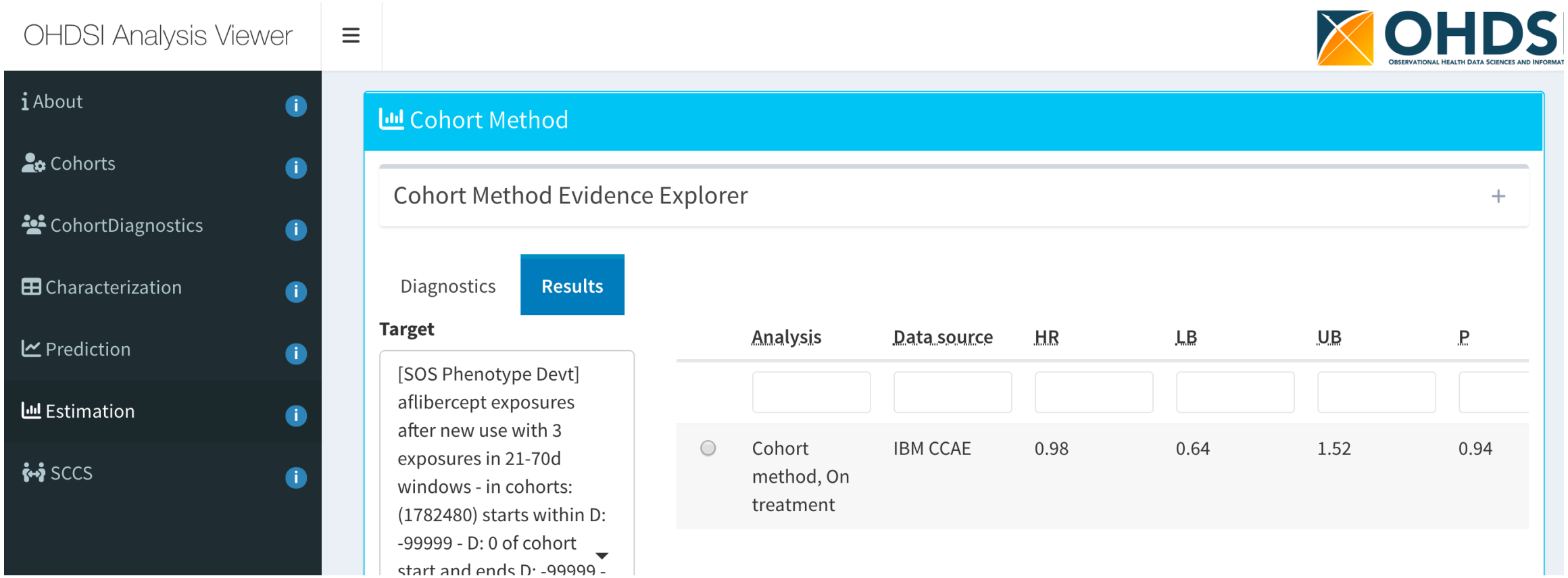
Characterization

Explore incidence rates, event timing, compare those target patients with and without the outcome during the time at risk...



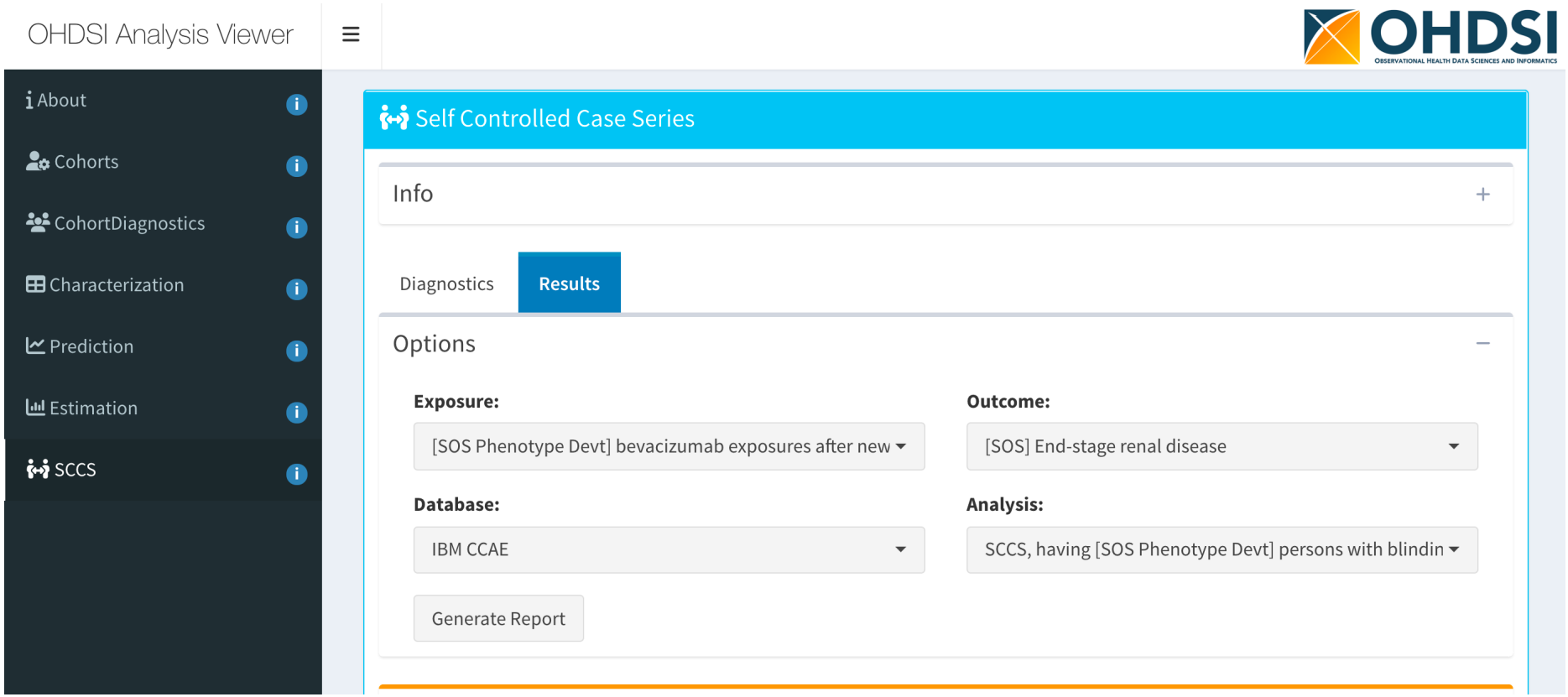
Cohort Method

View diagnostics and cohort method results (if they passed diagnostics), explore the analyses...



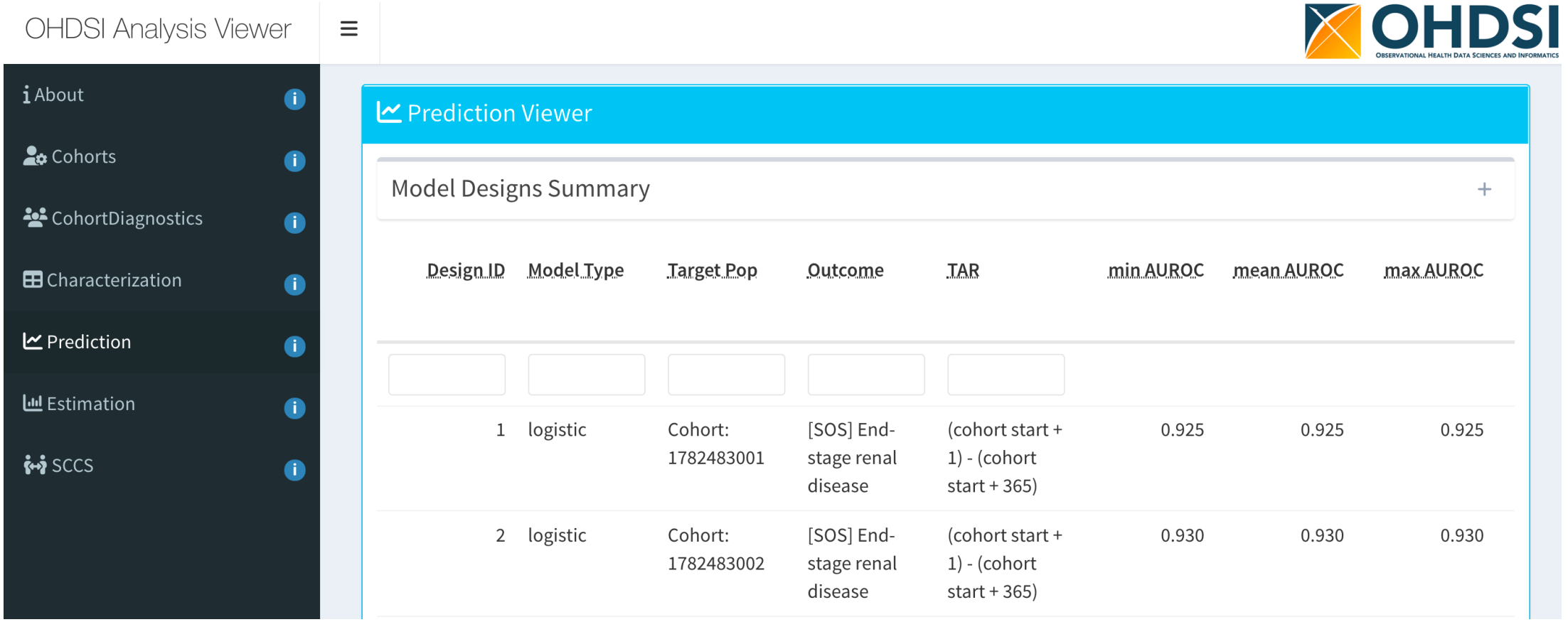
Self-Controlled Case Series

View diagnostics and SCCS method results (if they passed diagnostics), explore the analyses...



PatientLevelPrediction

View how feasible it was to predict the outcomes within the target populations during the time at risk...



Design ID	Model Type	Target Pop	Outcome	TAR	min AUROC	mean AUROC	max AUROC
1	logistic	Cohort: 1782483001	[SOS] End-stage renal disease	(cohort start + 1) - (cohort start + 365)	0.925	0.925	0.925
2	logistic	Cohort: 1782483002	[SOS] End-stage renal disease	(cohort start + 1) - (cohort start + 365)	0.930	0.930	0.930

We have a central shiny app

- Results will be uploaded to https://data.ohdsi.org/AntiVegfKidneyFailure/
- Check out that app throughout the week as more results are added...