

ImmPort Hands-on Training

Zicheng Hu
UCSF and ImmPort team

Outline

- Find studies related to anti-TNF α and rheumatoid arthritis.
- Download the study of interest (SDY824).
- Start Jupyter notebook server.
- Use Jupyter notebook for data preparation.
- Use Jupyter notebook for data visualization and statistical analysis.

Navigate to ImmPort Shared Data: <https://immpart.org/shared>

The screenshot shows the ImmPort Shared Data homepage. A red circle labeled '1' highlights the browser's address bar showing the URL immpart.org/shared/home. A red circle labeled '2' highlights the search bar at the top of the page containing the text "e.g. influenza (minimum 3 characters)".

1 To better understand the COVID-19 pandemic, ImmPort presents an opportunity to leverage legacy studies on infectious diseases, including Influenza (over 100 studies) and other respiratory-like illnesses from diverse cohorts (e.g., age, race, gender) sourced from NIAID-sponsored programs and beyond. Please contact the Helpdesk (ImmPort_Helpdesk@immpart.org) with questions regarding the shared data sets or data uploads.

Check out the ImmPort links:

- COVID-19 studies
- Influenza studies
- Respiratory-like illnesses studies
- Viral infectious diseases studies

IMMPORT Shared Data
Your site for searching and downloading shared data

Announcements
September 30, 2020 - ImmPort Data Release 36 is out with 10 new studies. Find more information in the [Data Release Notes](#).

Home Search e.g. influenza (minimum 3 characters)

Data Summary: Release 36, September 2020

Click on the counts with icon to visualize the count breakdown

Studies 445	Subjects 62,296	Diseases 112
Experiments 1,797	Total Results 5,933,719	Lab Tests 1,208,133

Bubble Summary: Research Focus by Assay Type

Click on a bubble to zoom in

Research Focus by Assay Type

Matrix Summary - Assay by Assay

Search for studies related to rheumatoid arthritis and TNF

Home Search TNF 2 Reset

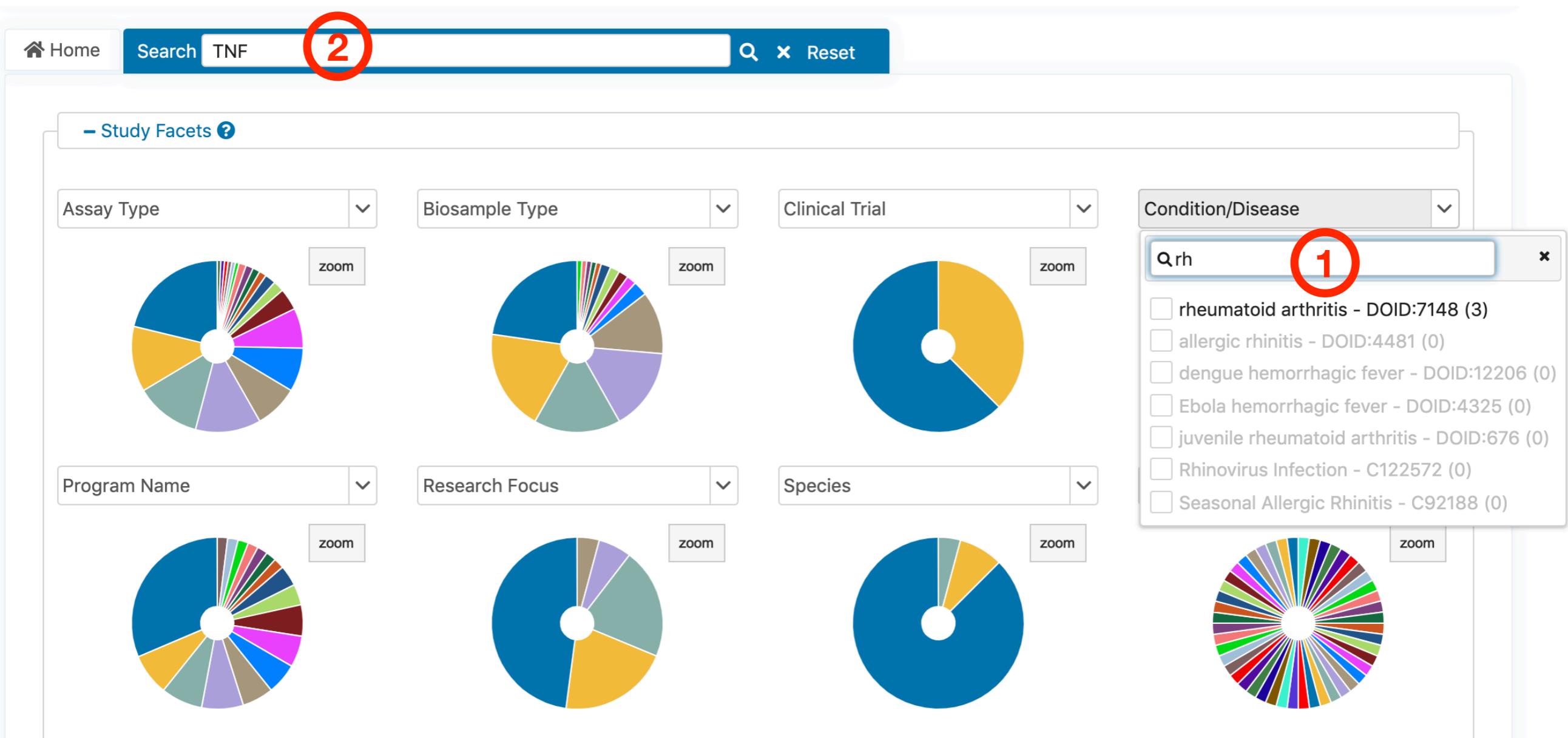
- Study Facets ?

Assay Type zoom Biosample Type zoom Clinical Trial zoom Condition/Disease zoom

Program Name zoom Research Focus zoom Species zoom

Condition/Disease search bar: rh 1 x

- rheumatoid arthritis - DOID:7148 (3)
- allergic rhinitis - DOID:4481 (0)
- dengue hemorrhagic fever - DOID:12206 (0)
- Ebola hemorrhagic fever - DOID:4325 (0)
- juvenile rheumatoid arthritis - DOID:676 (0)
- Rhinovirus Infection - C122572 (0)
- Seasonal Allergic Rhinitis - C92188 (0)



Choose SDY824

Found 3 studies in 12 ms

Click on the + or > icons to view match context when search term is specified

6 columns selectedDownload All StudiesExport

< < 1 > > 10 <

Showing 1 to 3 of 3

+	Study	Title	Pubmed Id	Research Focus	Latest Release Version	Latest Release Date
>	SDY146	Immune response evaluation in patients treated with TNF-alpha blockade (anti-TNF)	20298818 22177419	Immune Response	DR2	2013-06-12
1>	SDY824	Anti-TNF Agents in RA (ARA06)	18937634	Autoimmune	DR22	2017-06-16
>	SDY473	Lovastatin Therapy in Rheumatoid Arthritis (ARA02)		Autoimmune	DR15	2015-09-18

< < 1 > > 10 <

Showing 1 to 3 of 3

Summary of SDY824

SDY824 - Anti-TNF Agents in RA (ARA06) [!\[\]\(d0a1791f26d167e866e44ebbf83efebe_img.jpg\) Download](#)

Summary Design Adverse Event Assessment Interventions Medications Substance Demographics Lab Tests Mechanistic Assays Study Files

+ -

▼ Summary

Accession	SDY824
Title	Anti-TNF Agents in RA (ARA06)
DOI	10.21430/M3LU5YSBYO
Brief Description	The primary purpose of this study is to determine the effectiveness of two TNF inhibitors, etanercept and adalimumab, on memory B lymphocytes (B-cells) in the peripheral blood of participants with RA.
PI	Jennifer Anolik - University of Rochester School of Medicine
Type	
Condition Studied	rheumatoid arthritis
Start Date	2009-03-01
Detailed Description	This is a Phase IV, investigator-initiated, partially blinded, randomized, multi-center clinical trial designed to evaluate the mechanistic effects of TNF-a inhibition on clinical and mechanistic measures in RA patients. Subjects will be randomized to one of two active treatment arms, etanercept or adalimumab, and will receive treatment using standard dosing regimens for 24 weeks. Subjects will be randomized in a 2:1 ratio until 40 and 20 subjects are treated with etanercept and adalimumab, respectively. Clinical responses and serologic changes will be monitored throughout the study in all participants. Study visits are scheduled for Screening, Baseline/Treatment-Initiation, and at Weeks 12 and 24. In addition, to assess safety, blood draws are scheduled for Weeks 8 and 16, and phone calls are scheduled for Weeks 4, 8, 16, and 20. In order to assess peripheral blood B and T cell changes after treatment initiation, all participating subjects will have a blood sample analyzed by flow

View the study design

SDY824 - Anti-TNF Agents in RA (ARA06) [Download](#)

Summary **Design** Adverse Event Assessment Interventions Medications Substance Demographics Lab Tests Mechanistic Assays Study Files

1

+ -

▼ Arms or Cohorts

Accession	Name	Description
ARM3215	Etanercept Experimental	Participants receive a subcutaneous injection of etanercept once every week for 24 weeks
ARM3216	Adalimumab Experimental	Participants receive a subcutaneous injection of adalimumab once every 2 weeks for 24 weeks

View the available lab tests

SDY824 - Anti-TNF Agents in RA (ARA06)

[Download](#)

2

Summary

Design

Adverse Event

Assessment

Interventions

Medications

Substance

Demographics

Lab Tests

Mechanistic Assays

Study Files

+ -

▼ Lab Test Summary

Search Lab Test Summary

 Export

Lab Test Panel Name Preferred	Totals By	ARM3215	ARM3216
Blood Cell Count with Differential	biosample	292	127
Blood Cell Count with Differential	lab_test	4038	1764
Blood Cell Count with Differential	subject	43	20
Chemistry Test	biosample	133	65
Chemistry Test	biosample	79	38
Chemistry Test	lab_test	532	257
Chemistry Test	lab_test	315	152
Chemistry Test	subject	39	19
Chemistry Test	subject	43	20
Immunology Test	biosample	118	58

Outline

- Find studies related to anti-TNF α and rheumatoid arthritis.
- Download the study of interest (SDY824).
- Start Jupyter notebook server.
- Use Jupyter notebook for data preparation.
- Use Jupyter notebook for data visualization and statistical analysis.

Login for downloading data (skip the step during the online tutorial)



 Username

 Your username is not your email address

 Password

By checking the "**I Accept**" box below, you confirm that you have read and accept all the terms and conditions without limitation of the **User Agreement for the NIAID Immunology Database and Analysis Portal.**

I Accept

→ Login

Forgot Password?

Download the structured data (skip the step during the online tutorial)

* Archive files are disabled for download in this view, please navigate into the archive directory to download them.

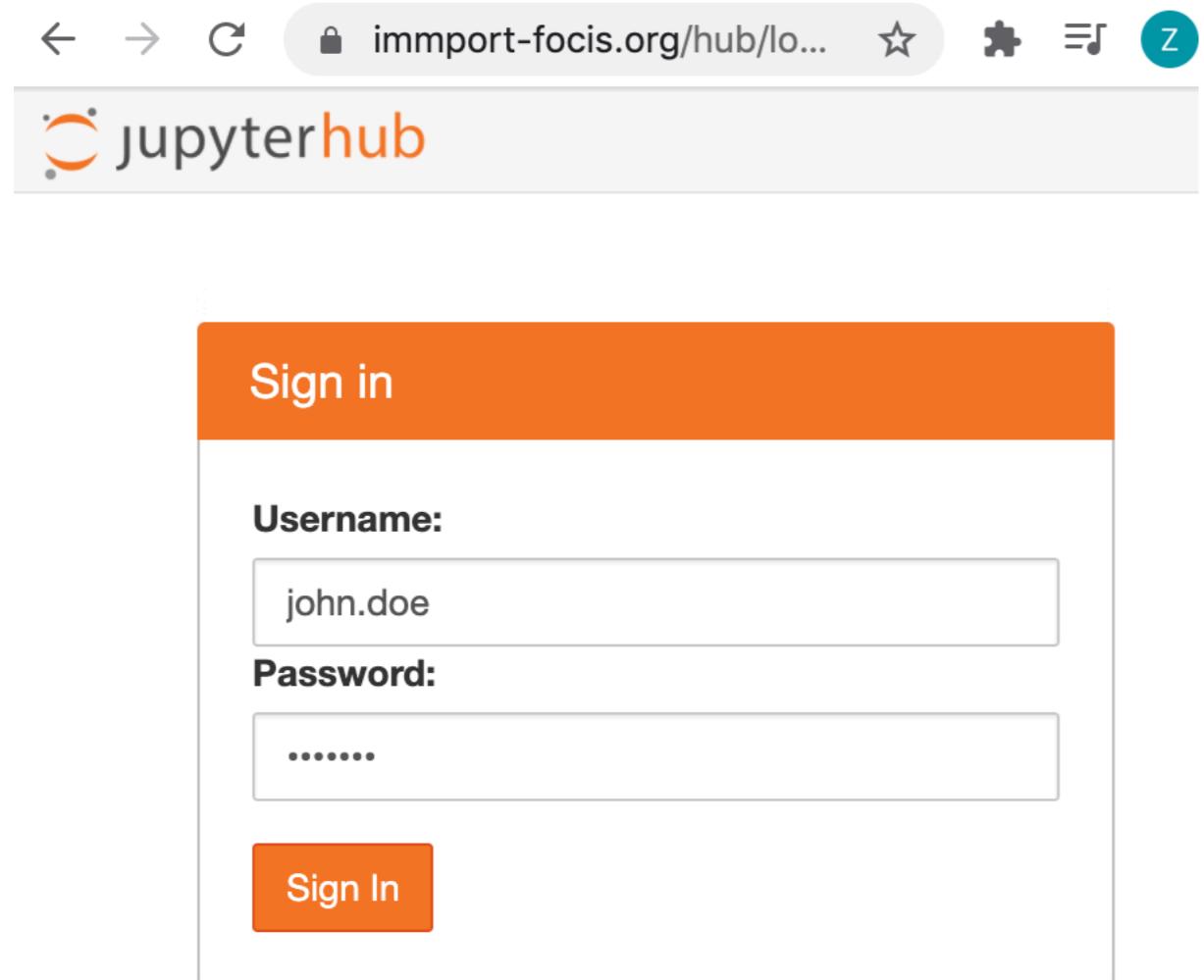
	Name	Size	Last Modified
<input type="checkbox"/>	Protocols (1 files)	1.10 MB	Jun 14, 2017 11:15 AM
<input type="checkbox"/>	ResultFiles (721 files)	34.30 GB	Dec 21, 2017 12:56 PM
<input type="checkbox"/>	SDY824-DR36_MySQL.zip	2.74 MB	Sep 30, 2020 5:08 AM
<input type="checkbox"/>	SDY824-DR36_Subject_2_Flow_cytometry_compensation_or_control.txt	5.88 MB	Sep 30, 2020 5:08 AM
<input type="checkbox"/>	SDY824-DR36_Subject_2_Flow_cytometry_result.txt	372.18 KB	Sep 30, 2020 5:08 AM
<input type="checkbox"/>	SDY824-DR36_Subject_2_Flow_cytometry_workspace.txt	363.79 KB	Sep 30, 2020 5:08 AM
<input checked="" type="checkbox"/>	SDY824-DR36_Tab.zip	2.41 MB	Sep 30, 2020 5:08 AM
<input type="checkbox"/>	SDY824-DR36_manifest.txt	57.77 KB	Sep 30, 2020 5:08 AM
<input type="checkbox"/>	SDY824-DR36_summary.txt	134.00 bytes	Sep 30, 2020 5:08 AM
<input type="checkbox"/>	SDY824-DR36_table_count.txt	2.25 KB	Sep 30, 2020 5:08 AM
<input type="checkbox"/>	StudyFiles (62 files)	9.34 MB	Jun 14, 2017 11:15 AM
<input type="checkbox"/>	* archive (109 files)	157.52 MB	Sep 30, 2020 5:05 AM

Showing 1 to 12 of 12 records

Outline

- Find studies related to anti-TNF α and rheumatoid arthritis.
- Download the study of interest (SDY824).
- Start Jupyter notebook server.
- Use Jupyter notebook for data preparation.
- Use Jupyter notebook for data visualization and statistical analysis.

Navigate to the jupyter hub server: import-focus.org



User name:

lower case email address
before the @
(e.g. john.doe in the
john.doe@email.com)

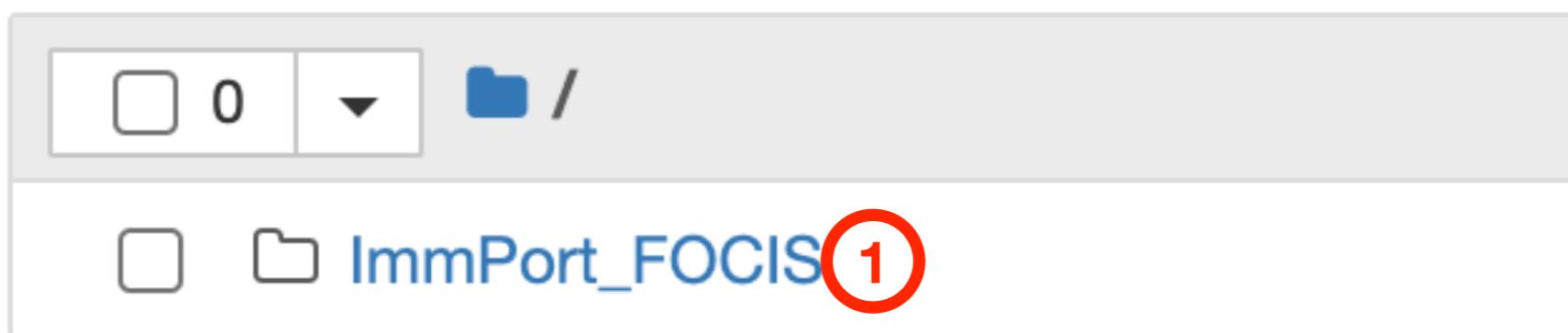
Password: import

Open the tutorial folder

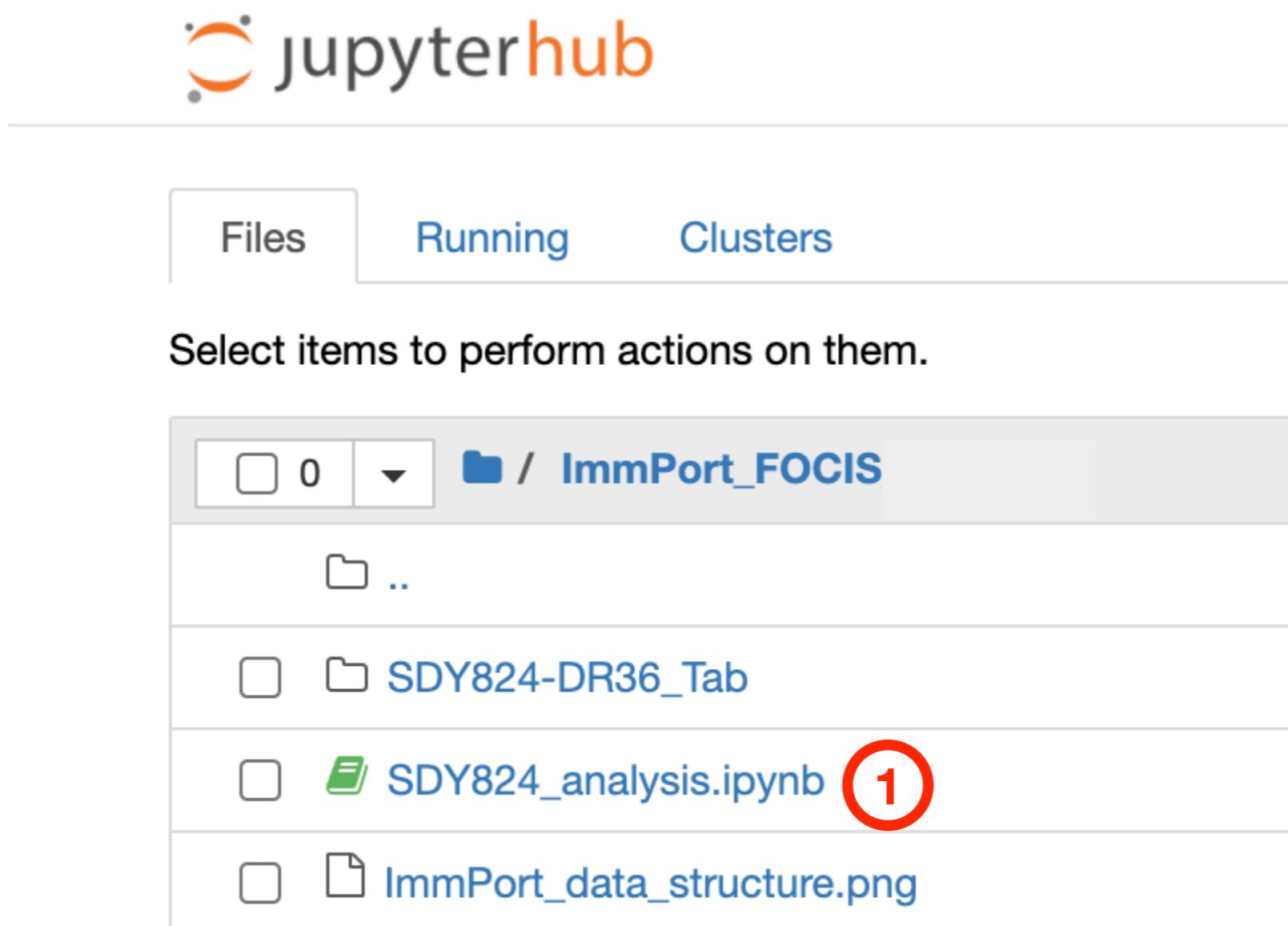


Files Running Clusters

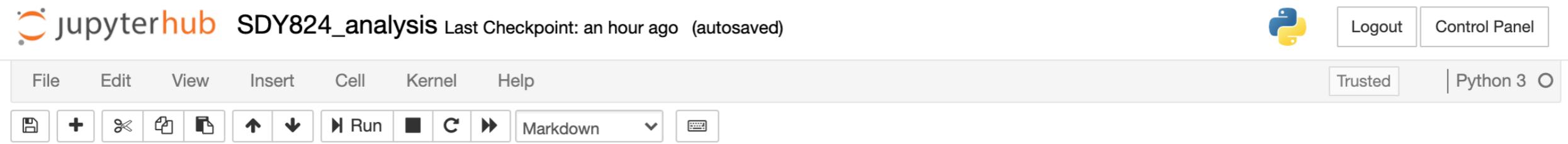
Select items to perform actions on them.



Open the SDY824_analysis.ipynb



Open the notebook: SDY824_analysis.ipynb



Exploratory Analysis of a Clinical trial on Anti-TNF Agents in Rheumatoid Arthritis Patients

In this tutorial, we will use python and jupyter notebook to analyze study [SDY824](#), a clinical trial that compared two anti-TNF drugs, Etanercept and Adalimumab, in Rheumatoid Arthritis Patients.

We will learn the following skills in the tutorial:

- Use python and the popular library Pandas to manipulate data.
- Navigate through the ImmPort database.
- Data visualization using the seaborn library.
- Perform statistical test using mixed effect models.

Using these skills, we will achieve the following scientific goals:

- Compare the demographics between the two treatment arms.
- Visualize the data from different lab tests.
- Identify measurements that are significantly different between the two treatment arms.

Introduction to the tutorial

Exploratory Analysis of a Clinical trial on Anti-TNF Agents in Rheumatoid Arthritis Patients

In this tutorial, we will use python and jupyter notebook to analyze study [SDY824](#), a clinical trial that compared two anti-TNF drugs, Etanercept and Adalimumab, in Rheumatoid Arthritis Patients.

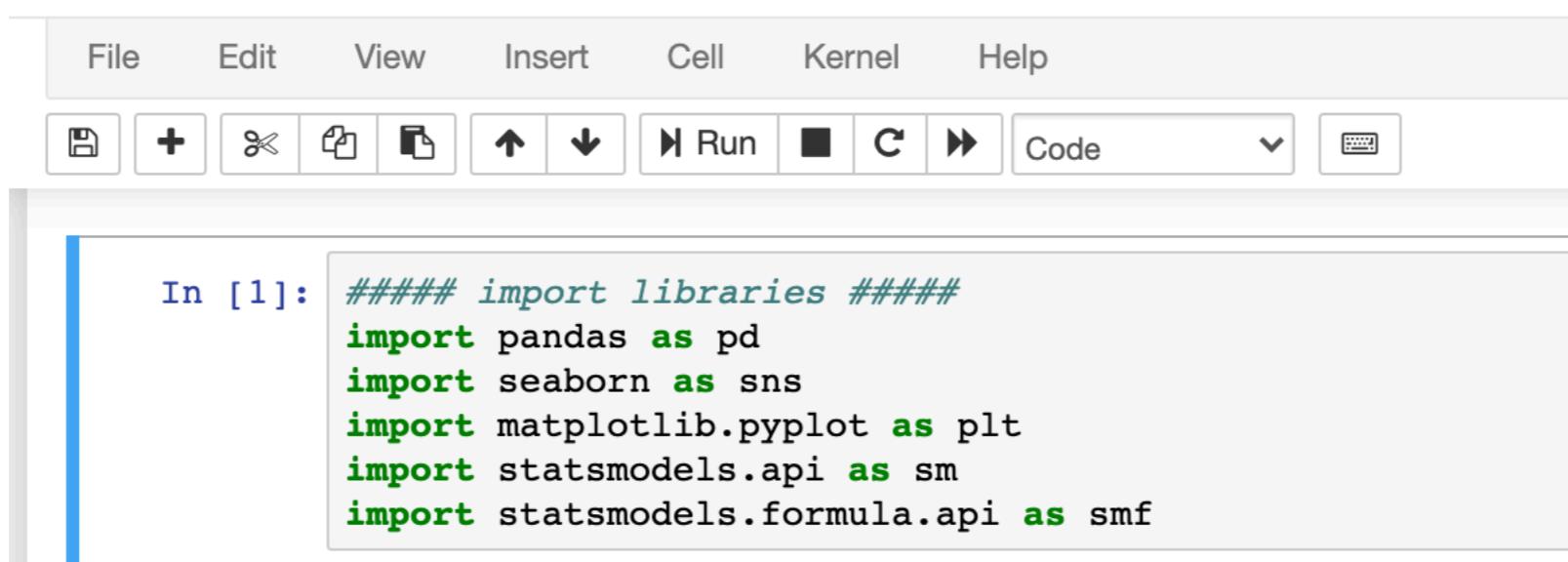
We will learn the following skills in the tutorial:

- Use python and the popular library Pandas to manipulate data.
- Navigate through the ImmPort database.
- Data visualization using the seaborn library.
- Perform statistical test using mixed effect models.

Using these skills, we will achieve the following scientific goals:

- Compare the demographics between the two treatment arms.
- Visualize the data from different lab tests.
- Identify measurements that are significantly different between the two treatment arms.

Import necessary libraries



The image shows a screenshot of a Jupyter Notebook interface. At the top, there is a menu bar with options: File, Edit, View, Insert, Cell, Kernel, and Help. Below the menu bar is a toolbar with various icons for file operations like saving, opening, and deleting, as well as buttons for running cells, kernel selection, and code completion. The main area is a code cell labeled "In [1]". The code in the cell is as follows:

```
In [1]: ##### import libraries #####
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import statsmodels.api as sm
import statsmodels.formula.api as smf
```

View the design of the study

```
In [2]: ##### view the design of the study #####
arms = pd.read_csv("SDY824-DR36_Tab/Tab/arm_or_cohort.txt",sep='\t')
arms = arms.query('STUDY_ACCESSION == "SDY824"')
display(arms.head())
print(arms.DESCRIPTION.values)
```

ARM_ACCESSION	DESCRIPTION	NAME	STUDY_ACCESSION	TYPE	WORKSPACE_ID
0	ARM3215 Participants receive a subcutaneous injection ...	Etanercept Experimental	SDY824	Experimental	5300
1	ARM3216 Participants receive a subcutaneous injection ...	Adalimumab Experimental	SDY824	Experimental	5300

```
['Participants receive a subcutaneous injection of etanercept once every week for 24 weeks'
 'Participants receive a subcutaneous injection of adalimumab once every 2 weeks for 24 weeks']
```

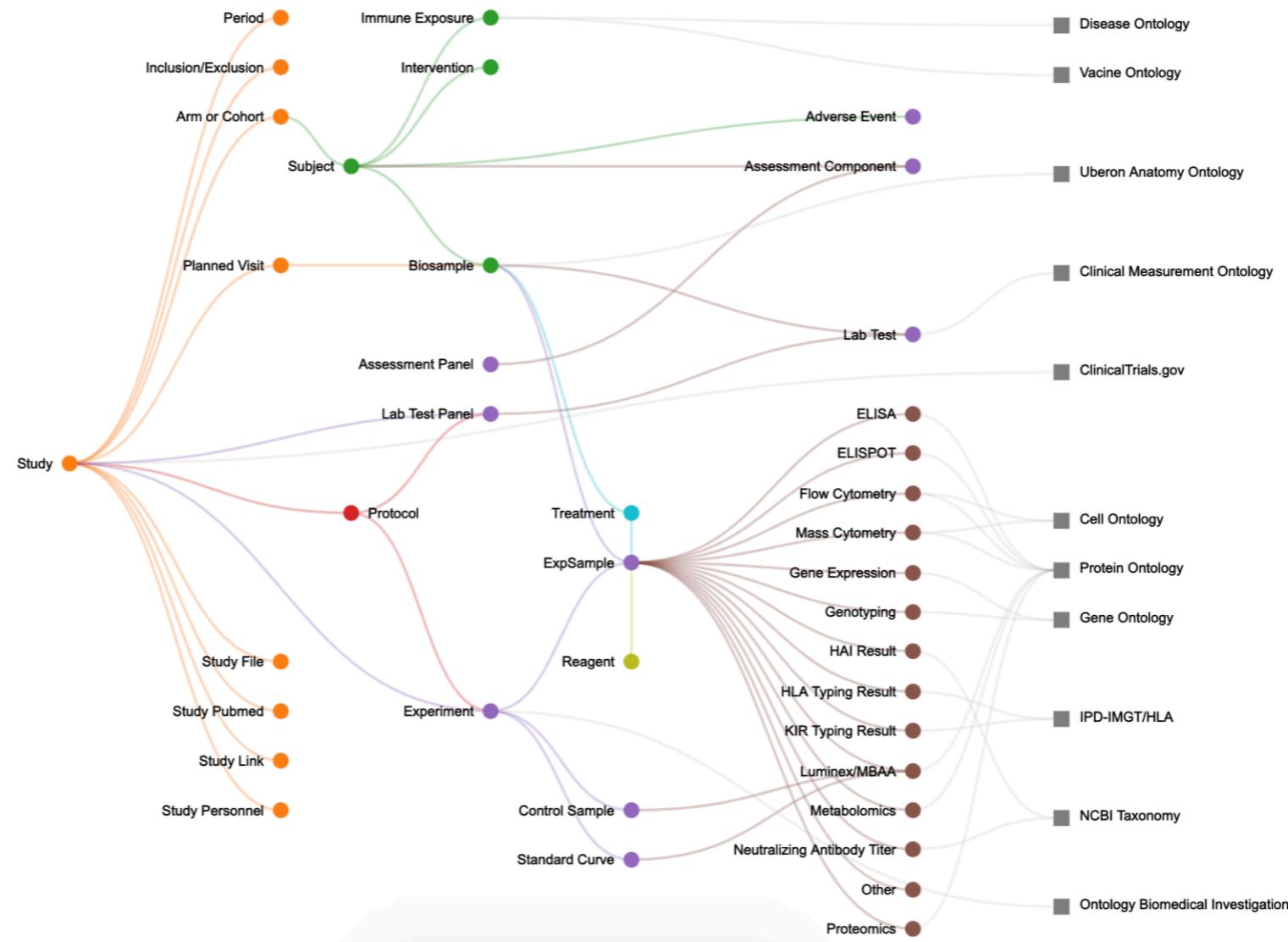
Outline

- Find studies related to anti-TNF α and rheumatoid arthritis.
- Download the study of interest (SDY824).
- Start Jupyter notebook server.
- Use Jupyter notebook for data preparation.
- Use Jupyter notebook for data visualization and statistical analysis.

ImmPort data model

Navigate the ImmPort data model

In the next few code blocks, we will navigate through the ImmPort data model to gather subject information and lab tests measurement.



Assign subject to the arms

```
In [3]: ##### assign subject to the arms #####
arm2subj = pd.read_csv("SDY824-DR36_Tab/Tab/arm_2_subject.txt",sep='\t')
subj = arm2subj.merge(arms, how='inner', on="ARM_ACCESSION")
subj = subj[["ARM_ACCESSION", "SUBJECT_ACCESSION", "MIN SUBJECT AGE", "NAME", "STUDY_ACCESSION"]]
display(subj.head())
```

	ARM_ACCESSION	SUBJECT_ACCESSION	MIN SUBJECT AGE	NAME	STUDY_ACCESSION
0	ARM3215	SUB169152	52	Etanercept Experimental	SDY824
1	ARM3215	SUB169153	43	Etanercept Experimental	SDY824
2	ARM3215	SUB169154	41	Etanercept Experimental	SDY824
3	ARM3215	SUB169156	32	Etanercept Experimental	SDY824
4	ARM3215	SUB169157	45	Etanercept Experimental	SDY824

Join with subject information

```
#### join with subject information #####
subj_info = pd.read_csv("SDY824-DR36_Tab/Tab/subject.txt",sep='\t')
subj_info = subj_info[["SUBJECT_ACCESSION", "GENDER", "RACE"]]
subj_info = subj_info.merge(subj, how='inner', on=["SUBJECT_ACCESSION"])
display(subj_info.head())
```

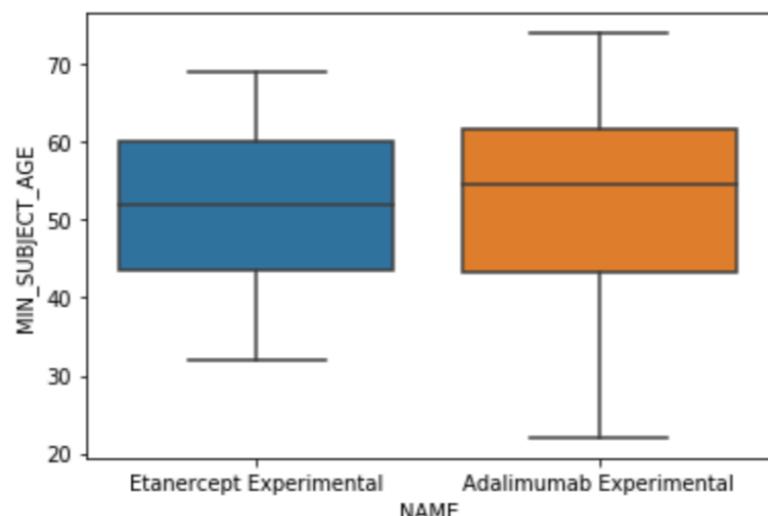
	SUBJECT_ACCESSION	GENDER	RACE	ARM_ACCESSION	MIN SUBJECT AGE	NAME	STUDY_ACCESSION
0	SUB169152	Female	White	ARM3215	52	Etanercept Experimental	SDY824
1	SUB169153	Female	White	ARM3215	43	Etanercept Experimental	SDY824
2	SUB169154	Female	White	ARM3215	41	Etanercept Experimental	SDY824
3	SUB169155	Female	White	ARM3216	64	Adalimumab Experimental	SDY824
4	SUB169156	Male	Black or African American	ARM3215	32	Etanercept Experimental	SDY824

Plot demographics

```
In [5]: ##### plot demographics #####
sns.boxplot(x="NAME", y="MIN SUBJECT AGE", data=subj_info)
plt.show()

gender_plot = subj_info.groupby(['NAME'])['GENDER'].value_counts(normalize=True)
gender_plot = gender_plot.rename('proportion').reset_index()
sns.barplot(x="GENDER", y="proportion", hue='NAME', data=gender_plot)
plt.show()

race_plot = subj_info.groupby(['NAME'])['RACE'].value_counts(normalize=True)
race_plot = race_plot.rename('proportion').reset_index()
sns.barplot(y="RACE", x="proportion", hue='NAME', data=race_plot)
plt.show()
```



Map biosample to the subjects

```
In [6]: ##### assign biosample to the subjects #####
biosample = pd.read_csv("SDY824-DR36_Tab/Tab/biosample.txt",sep='\t')
biosample = biosample[["BIOSAMPLE_ACCESSION", "STUDY_TIME_COLLECTED", "SUBJECT_ACCESSION", "STUDY_ACCESSION"]]
biosample = biosample.merge(subj_info, how='inner', on=["SUBJECT_ACCESSION", "STUDY_ACCESSION"])
display(biosample.head())
```

	BIOSAMPLE_ACCESSION	STUDY_TIME_COLLECTED	SUBJECT_ACCESSION	STUDY_ACCESSION	GENDER	RACE	ARM_ACCESSION	MIN SUBJECT AGE
0	BS913115	66	SUB169158	SDY824	Female	White	ARM3216	61
1	BS913116	122	SUB169158	SDY824	Female	White	ARM3216	61
2	BS913117	182	SUB169158	SDY824	Female	White	ARM3216	61
3	BS913315	-7	SUB169158	SDY824	Female	White	ARM3216	61
4	BS913316	13	SUB169158	SDY824	Female	White	ARM3216	61

Add lab test results to the table

```
In [7]: ##### join with lab test result #####
lab = pd.read_csv("SDY824-DR36_Tab/Tab/lab_test.txt",sep='\t')
lab = lab[['BIOSAMPLE_ACCESSION','NAME_PREFERRED','RESULT_UNIT_REPORTED','RESULT_VALUE_PREFERRED']]
lab = lab.merge(biosample, how='inner', on=['BIOSAMPLE_ACCESSION'])
lab = lab.dropna()
lab['name_plus_unit'] = lab['NAME_PREFERRED'] + " in " + lab['RESULT_UNIT_REPORTED']
display(lab.head())
```

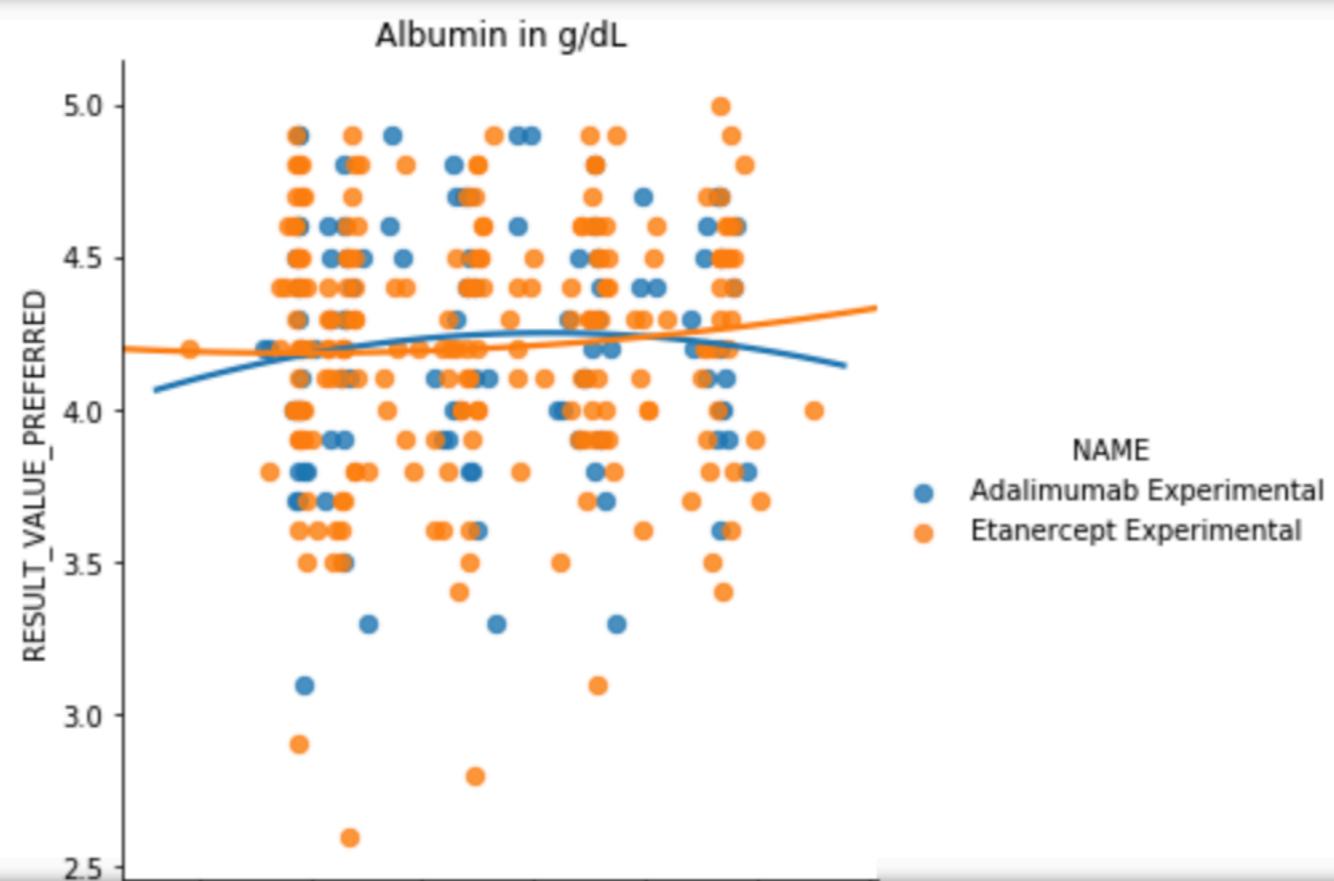
	BIOSAMPLE_ACCESSION	NAME_PREFERRED	RESULT_UNIT_REPORTED	RESULT_VALUE_PREFERRED	STUDY_TIME_COLLECTED	SUBJECT_ACCESSION	ST
0	BS913115	Albumin	g/dL	4.0	66	SUB169158	
1	BS913115	Alanine Aminotransferase	U/L	13.0	66	SUB169158	
2	BS913115	Aspartate Aminotransferase	U/L	20.0	66	SUB169158	
3	BS913115	Creatinine	mg/dL	0.8	66	SUB169158	
4	BS913115	Hematocrit	%	36.3	66	SUB169158	

Outline

- Find studies related to anti-TNF α and rheumatoid arthritis.
- Download the study of interest (SDY824).
- Start Jupyter notebook server.
- Use Jupyter notebook for data preparation.
- Use Jupyter notebook for data visualization and statistical analysis.

Visualize the lab test results

```
In [8]: ##### visualize the lab test results #####
for name in lab["name_plus_unit"].unique():
    lab_data = lab[lab["name_plus_unit"] == name]
    sns.lmplot(x="STUDY_TIME_COLLECTED", y="RESULT_VALUE_PREFERRED",
                order=2, ci=None, scatter_kws={"s": 40},
                hue="NAME", data=lab_data)
    plt.title(name)
    plt.show()
```



Convert time to categorical variable

```
In [9]: ##### convert time to catogorical variable #####
lab['time'] = ''
lab.loc[lab.STUDY_TIME_COLLECTED<=0, "time"] = "T0"
lab.loc[(lab.STUDY_TIME_COLLECTED>0) & (lab.STUDY_TIME_COLLECTED<=50), 'time'] = "T1"
lab.loc[(lab.STUDY_TIME_COLLECTED>50) & (lab.STUDY_TIME_COLLECTED<=100), 'time'] = "T2"
lab.loc[(lab.STUDY_TIME_COLLECTED>100) & (lab.STUDY_TIME_COLLECTED<=150), 'time'] = "T3"
lab.loc[(lab.STUDY_TIME_COLLECTED>150), 'time'] = "T4"
lab.head()
```

Analyze the lab tests using mixed models

```
In [13]: ##### analyze lab tests using mixed models #####
test_result = pd.DataFrame({'lab':lab["name_plus_unit"].unique(), 'p':1})
for i in range(test_result.shape[0]):
    lab_data = lab[lab["name_plus_unit"] == test_result.lab[i]]
    if(lab_data.shape[0]<20): continue
    mix_model = smf.mixedlm("RESULT_VALUE_PREFERRED ~ time + NAME", lab_data, groups=lab_data["SUBJECT_ACCESSION"])
    mix_model = mix_model.fit()
    test_result.loc[i,"p"] = mix_model.pvalues["NAME[T.Etanercept Experimental]"]
test_result = test_result.sort_values("p").reset_index(drop = True)
display(test_result)

/Users/zichenghu/anaconda3/lib/python3.7/site-packages/statsmodels/regression/mixed_linear_model.py:2094: ConvergenceWarning
  Warning: The MLE may be on the boundary of the parameter space.
  warnings.warn(msg, ConvergenceWarning)
```

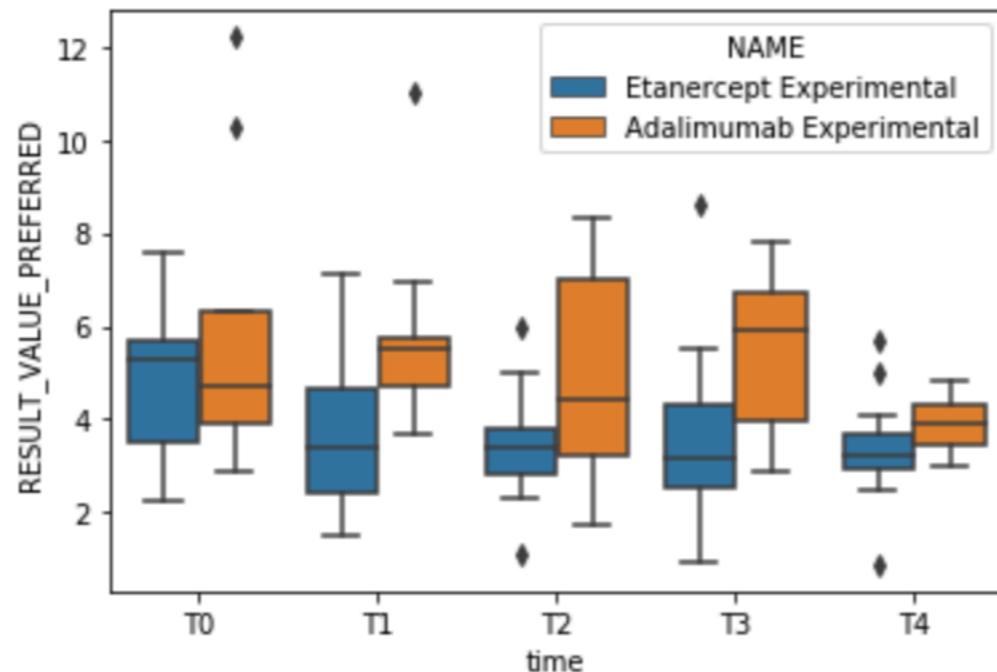
	lab	p
0	Neutrophils in 1000/Microliter	0.037393
1	Eosinophils in 1000/Microliter	0.068776
2	Aspartate Aminotransferase in U/L	0.172889
3	Platelets in 10^3/uL	0.213686
4	Basophils in Percent	0.216790
5	Monocytes in Percent	0.282496
6	Alanine Aminotransferase in U/L	0.304832
7	Ery. Mean Corpuscular HGB Concentration in g/dL	0.305237
8	Ery. Mean Corpuscular Hemoglobin in pg	0.319757

Visualize neutrophil number using a box plot

```
In [11]: ##### look at neutrophil numbers again #####
```

```
neutrophil_lab = lab[lab["name_plus_unit"] == "Neutrophils in 1000/Microliter"]
neutrophil_lab = neutrophil_lab.sort_values('time')
sns.boxplot(x="time", y="RESULT_VALUE_PREFERRED",
             hue="NAME", data=neutrophil_lab)
```

```
Out[11]: <matplotlib.axes._subplots.AxesSubplot at 0x1c23c63cd0>
```



Summary

In this tutorial, we analyzed the data from SDY 824 and achieved the following:

- Navigated through the ImmPort data model and gathered subject-level information and measurements from lab tests.
- We confirmed that the cohorts in the two arms are comparable.
- We visualized the lab test data and used mixed linear models to identify differences between the two arms.
- The result shows that there are a higher number of neutrophils in the adalimumab arm.