APACHE ZEPPELIN TRAINING

A gentle intro with hands-on session

HELLO!

Ahyoung Ryu

A software engineer @NFLabs

A committer of Apache Zeppelin

Github @AhyoungRyu

HELLO!

Mina Lee

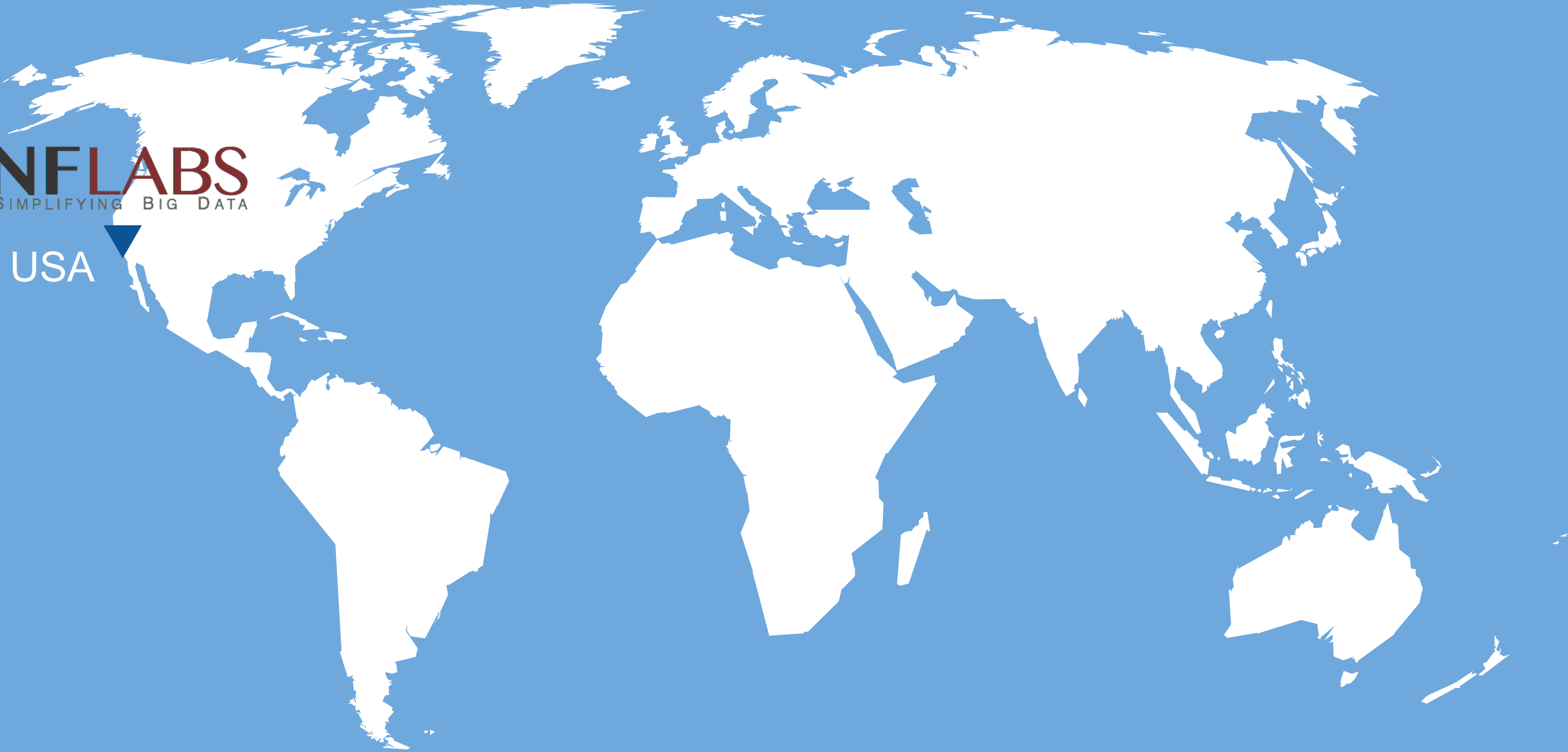
A software engineer @NFLabs

A PMC of Apache Zeppelin

Github @minahlee

Located

@ California, USA



NFLABS
SIMPLIFYING BIG DATA
NFLABS
SIMPLIFYING BIG DATA

Located

@ California, USA

Located

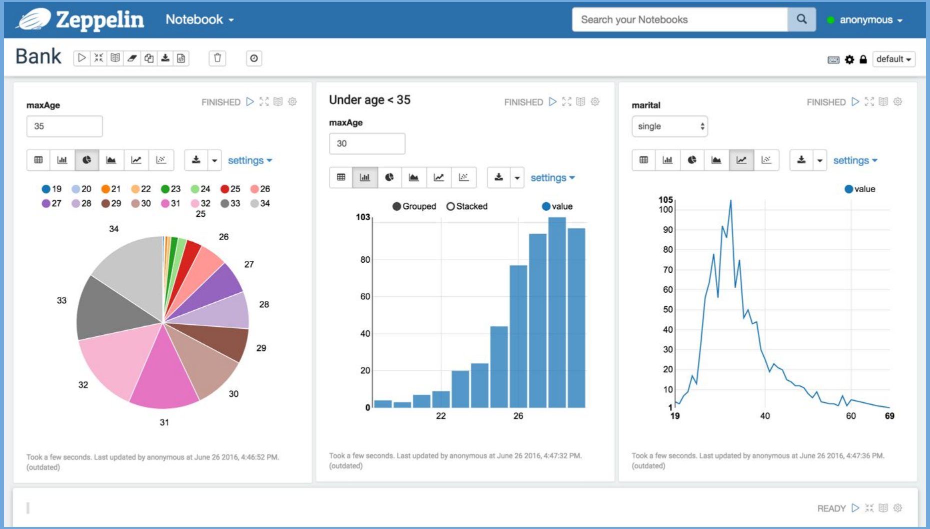
@ Seoul, Korea

"We're a creator of Apache Zeppelin!"

INDEX

Overview of Apache Zeppelin

Apache Zeppelin is used for...



Bank

maxAge: 35

FINISHED

Took a few seconds. Last updated by anonymous at June 26 2016, 4:46:52 PM. (outdated)

Under age < 35

maxAge: 30

FINISHED

Took a few seconds. Last updated by anonymous at June 26 2016, 4:47:32 PM. (outdated)

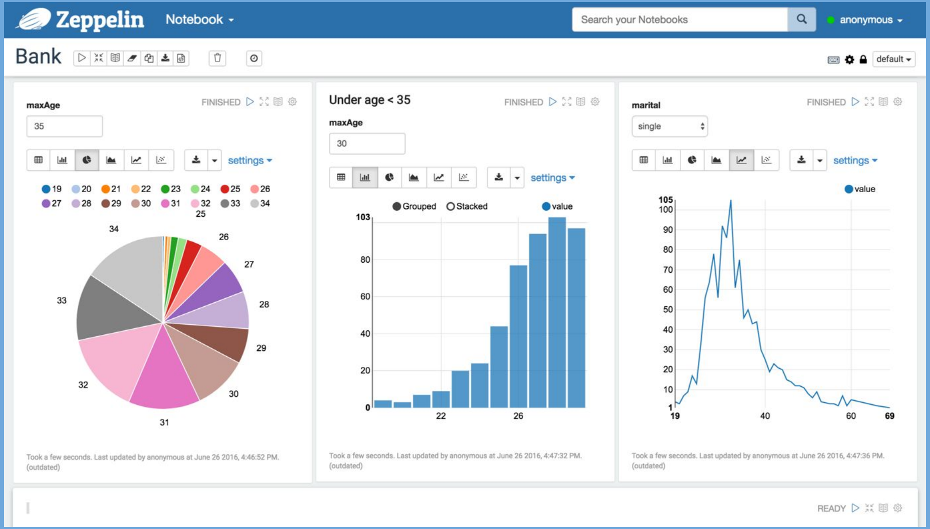
marital: single

FINISHED

Took a few seconds. Last updated by anonymous at June 26 2016, 4:47:36 PM. (outdated)

READY

Apache Zeppelin is used for...



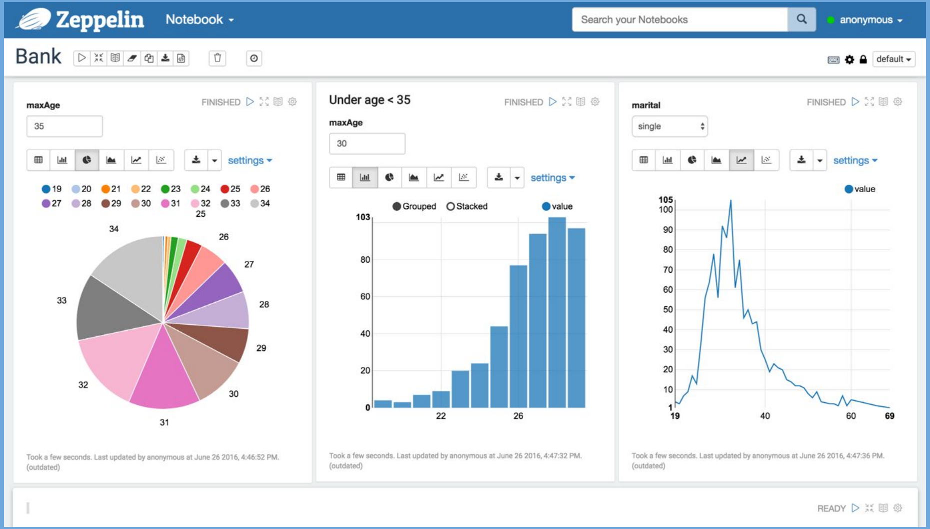
Took a few seconds. Last updated by anonymous at June 26 2016, 4:46:52 PM. (outdated)

Took a few seconds. Last updated by anonymous at June 26 2016, 4:47:32 PM. (outdated)

Took a few seconds. Last updated by anonymous at June 26 2016, 4:47:36 PM. (outdated)

- Data exploration and discovery
- Visualization
- Collaboration and publishing

Apache Zeppelin is used for...



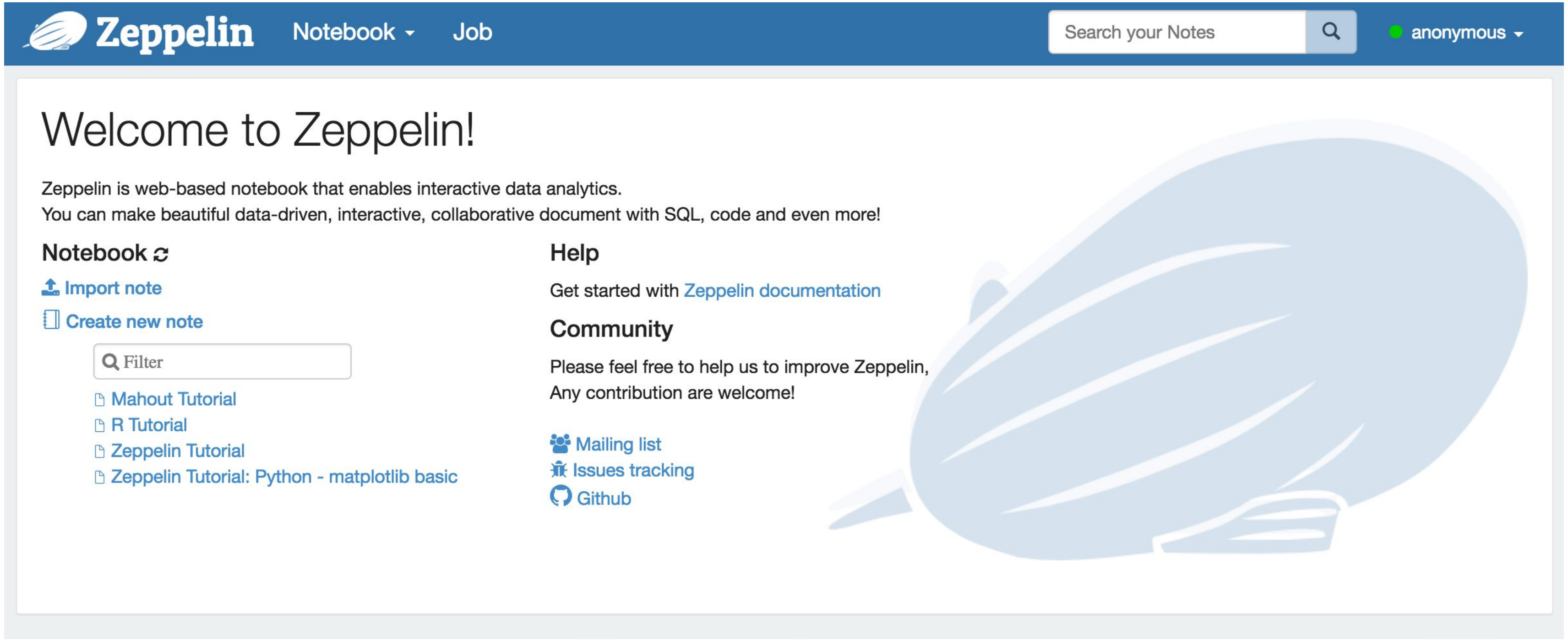
Took a few seconds. Last updated by anonymous at June 26 2016, 4:46:52 PM. (outdated)

Took a few seconds. Last updated by anonymous at June 26 2016, 4:47:32 PM. (outdated)

Took a few seconds. Last updated by anonymous at June 26 2016, 4:47:36 PM. (outdated)

- Data exploration and discovery
- Visualization
- Collaboration and publishing

“Modern Data Science Studio”



Welcome to Zeppelin!

Zeppelin is web-based notebook that enables interactive data analytics.
You can make beautiful data-driven, interactive, collaborative document with SQL, code and even more!

Notebook

Import note

Create new note

Filter

- Mahout Tutorial
- R Tutorial
- Zeppelin Tutorial
- Zeppelin Tutorial: Python - matplotlib basic

Help

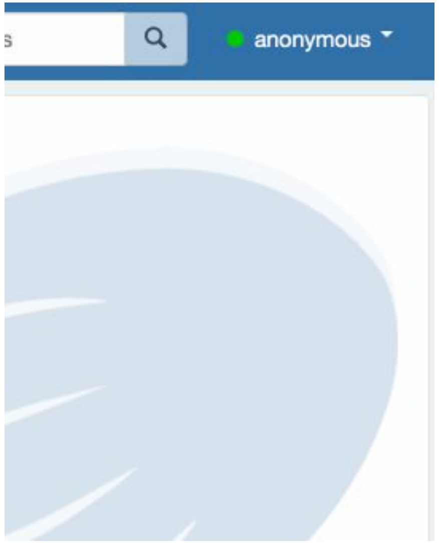
Get started with Zeppelin documentation

Community

Please feel free to help us to improve Zeppelin,
Any contribution are welcome!

- Mailing list
- Issues tracking
- Github

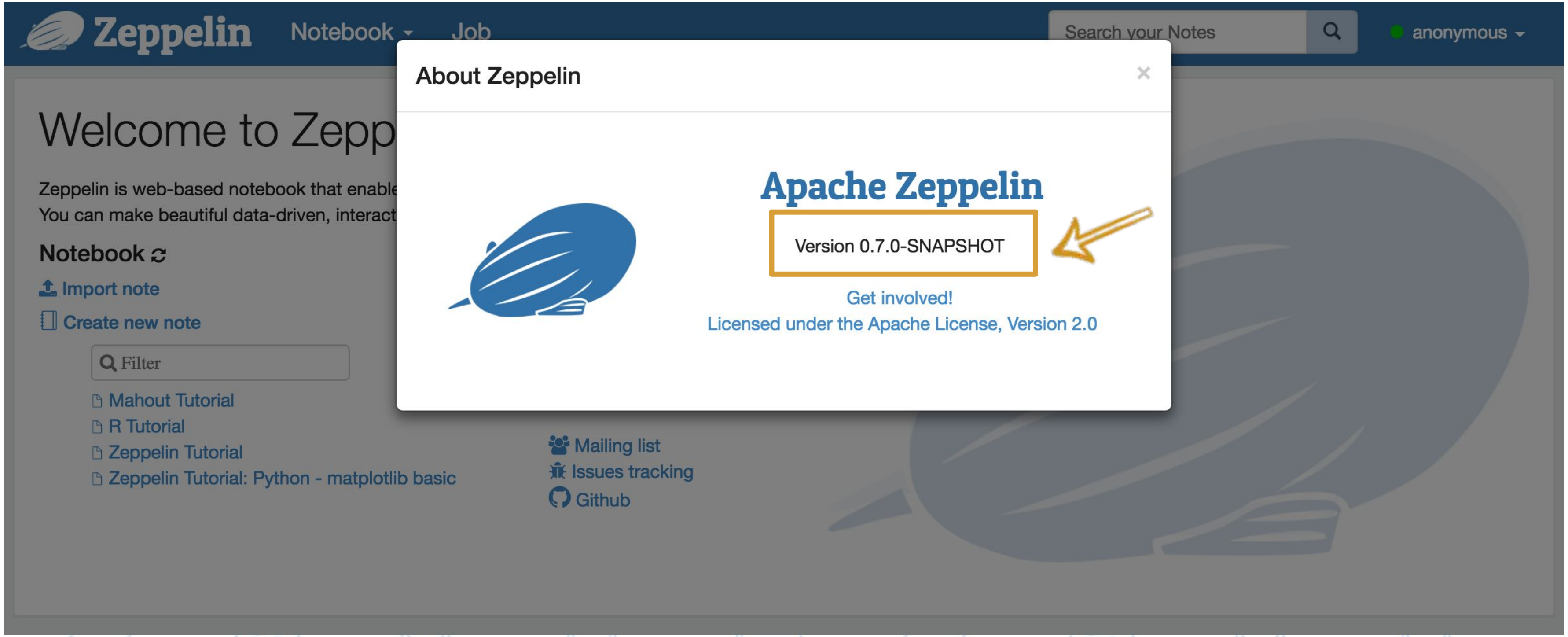
Main Menu



- About Zeppelin
- Interpreter Setting
- Notebook repos
- Credential
- Configuration

*Until you activate the **Shiro authentication**,*

*the username will be "**anonymous**"*



Welcome to Zepp

Zeppelin is web-based notebook that enable
You can make beautiful data-driven, interact

Notebook

Import note

Create new note

Filter

Mahout Tutorial

R Tutorial

Zeppelin Tutorial

Zeppelin Tutorial: Python - matplotlib basic

Mailing list

Issues tracking

Github

About Zeppelin

Apache Zeppelin

Version 0.7.0-SNAPSHOT

Get involved!

Licensed under the Apache License, Version 2.0

>_ shell

ALLUXIO Google BigQuery

cassandra elasticsearch Flink

APACHE GEODE APACHE HBASE hadoop HDFS HIVE

apache Ignite APACHE LENS

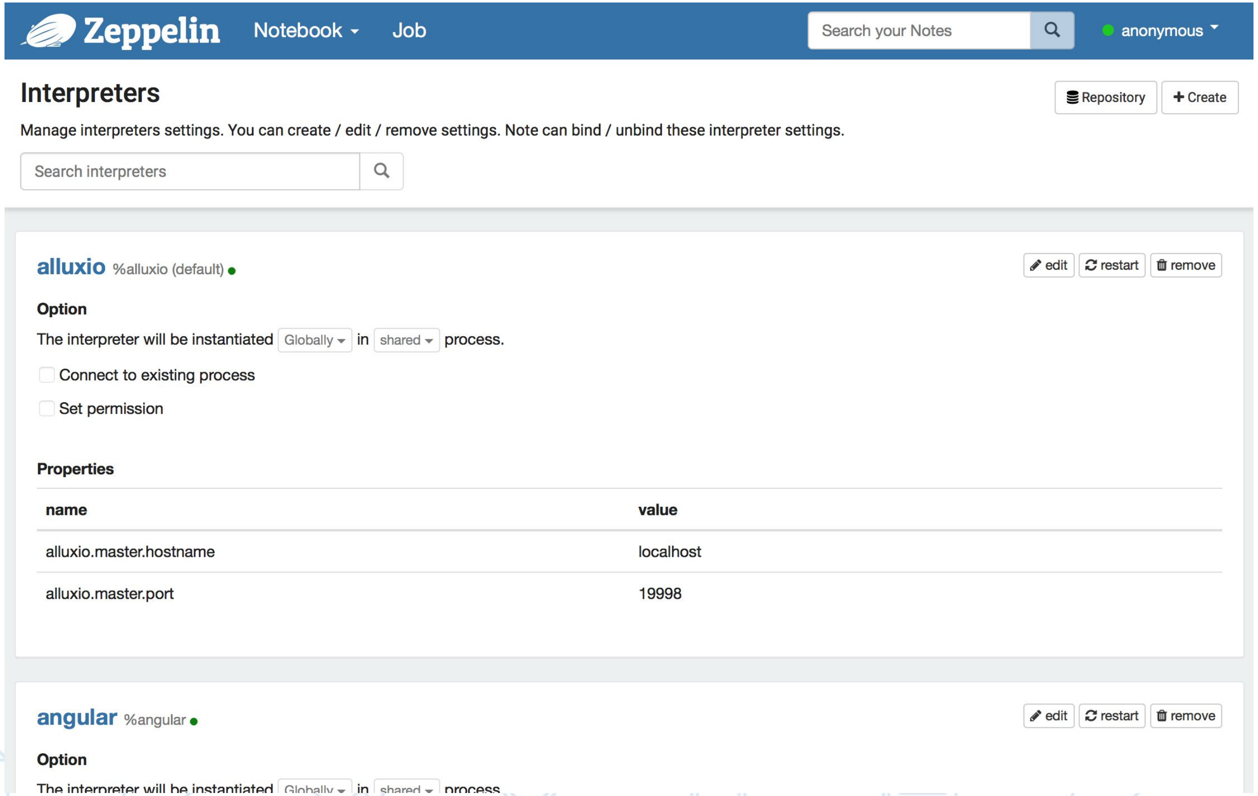
LIVY PostgreSQL python R

Scalding TAJO™

beam Powered by mahout **Markdown**

+20 ***interpreters*** *are available now!*

+ APACHE Spark™



Interpreters

Repository | Create

Manage interpreters settings. You can create / edit / remove settings. Note can bind / unbind these interpreter settings.

Search interpreters

alluxio %alluxio (default) ●

edit | restart | remove

Option

The interpreter will be instantiated Globally in shared process.

Connect to existing process

Set permission

Properties

name	value
alluxio.master.hostname	localhost
alluxio.master.port	19998

angular %angular ●

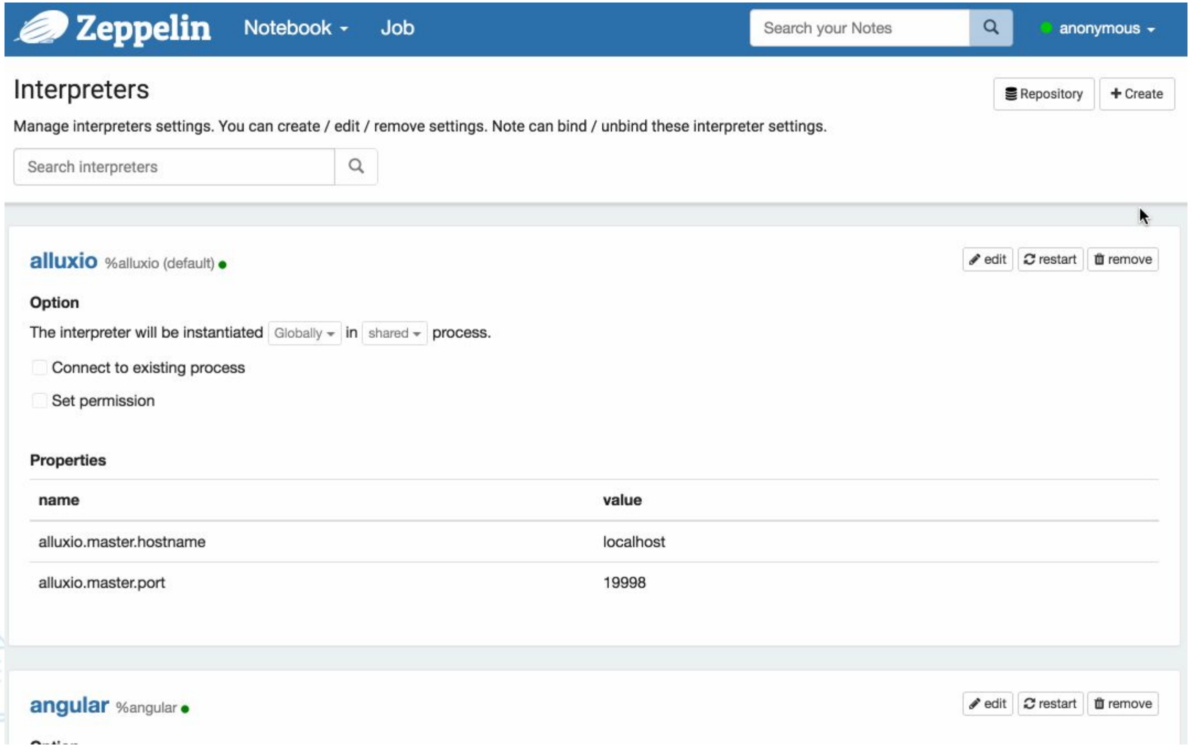
edit | restart | remove

Option

The interpreter will be instantiated Globally in shared process.

Interpreter Setting page

- Create/ edit/ remove
- Search
- Check repository info



Interpreters

Manage interpreters settings. You can create / edit / remove settings. Note can bind / unbind these interpreter settings.

alluxio %alluxio (default) ●

Option

The interpreter will be instantiated Globally in shared process.

Connect to existing process

Set permission

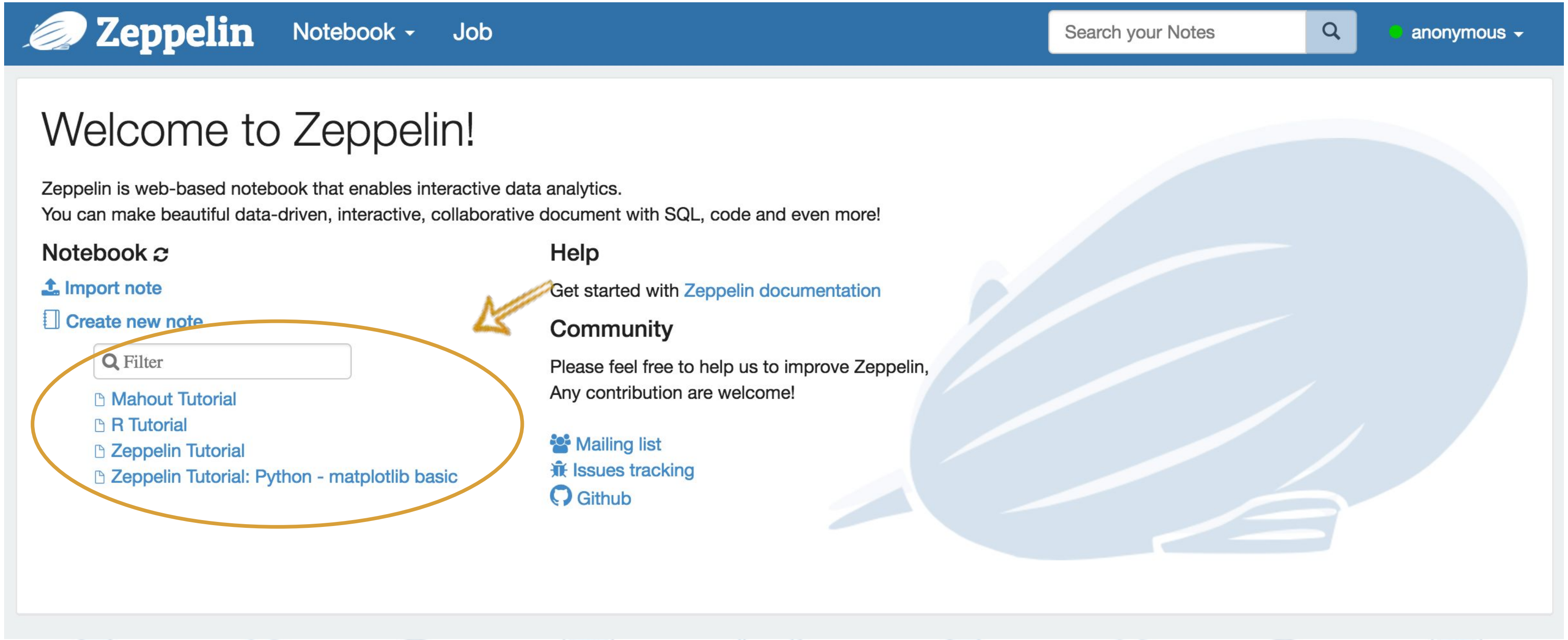
Properties

name	value
alluxio.master.hostname	localhost
alluxio.master.port	19998

angular %angular ●

Interpreter Creation

- Click “+Create” button
- Fill out the fields
- Default properties will be displayed
- Manage dependency if it’s needed



Zeppelin Notebook Job

Search your Notes

anonymous

Welcome to Zeppelin!

Zeppelin is web-based notebook that enables interactive data analytics.
You can make beautiful data-driven, interactive, collaborative document with SQL, code and even more!

Notebook

Import note

Create new note

Filter

- Mahout Tutorial
- R Tutorial
- Zeppelin Tutorial
- Zeppelin Tutorial: Python - matplotlib basic

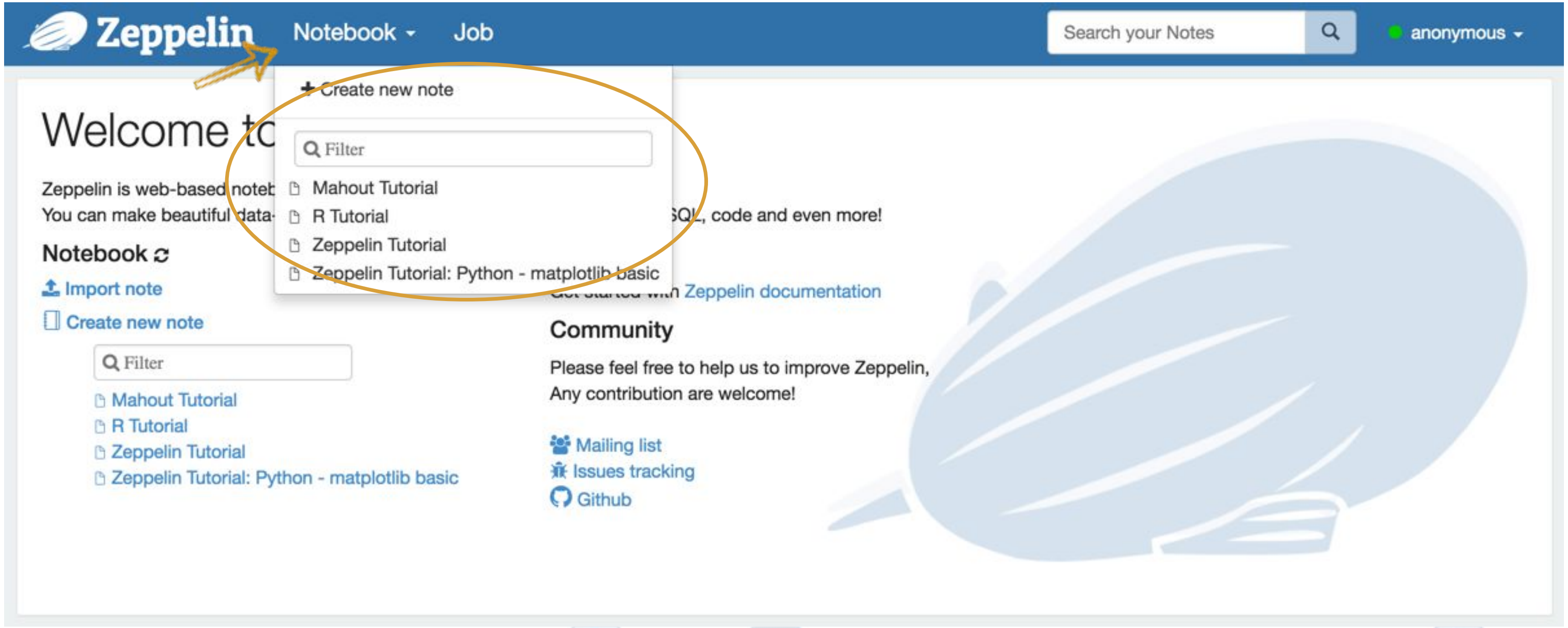
Help

Get started with Zeppelin documentation

Community

Please feel free to help us to improve Zeppelin,
Any contribution are welcome!

Mailing list
Issues tracking
Github



Zeppelin Notebook Job Search your Notes anonymous

+ Create new note

Filter

Mahout Tutorial

R Tutorial

Zeppelin Tutorial

Zeppelin Tutorial: Python - matplotlib basic

Welcome to

Zeppelin is web-based noteb

You can make beautiful data- SQL, code and even more!

Notebook

Import note

Create new note

Filter

Mahout Tutorial

R Tutorial

Zeppelin Tutorial

Zeppelin Tutorial: Python - matplotlib basic

Get started with Zeppelin documentation

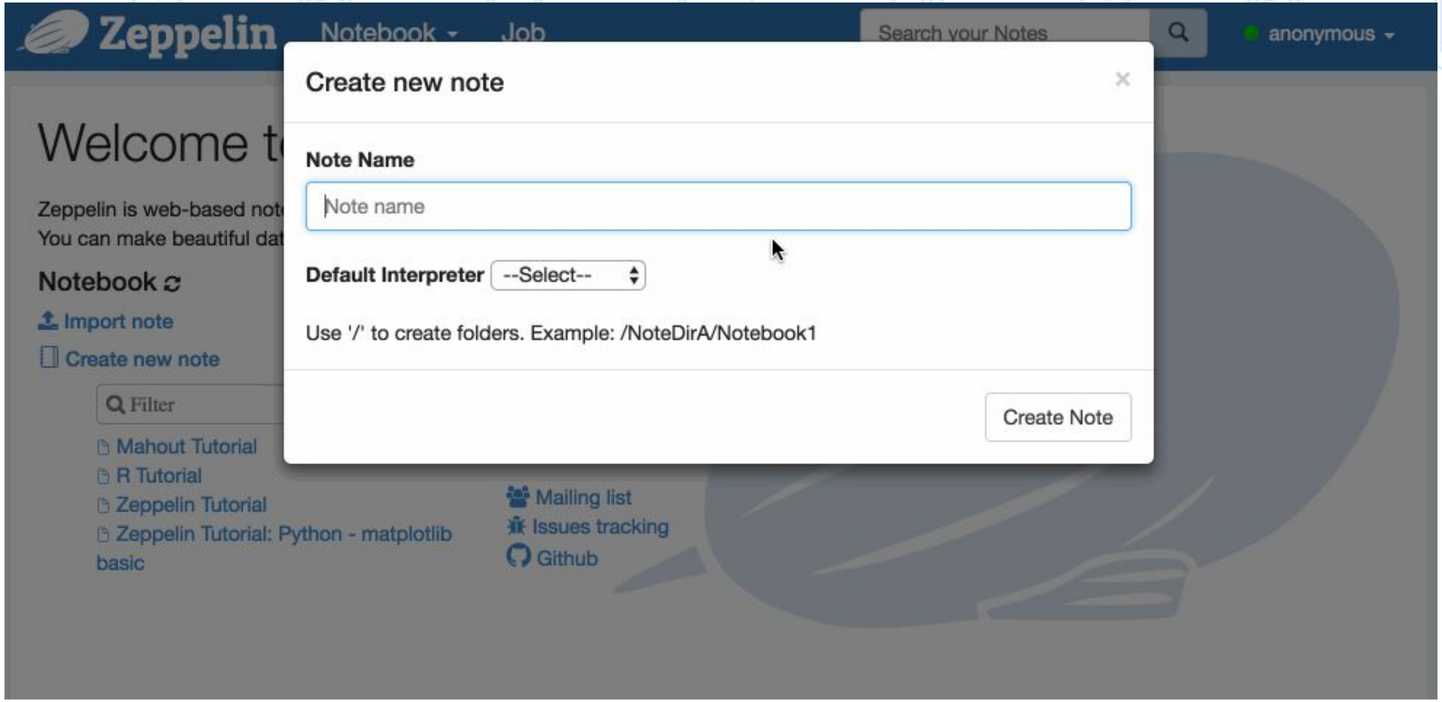
Community

Please feel free to help us to improve Zeppelin,
Any contribution are welcome!

Mailing list

Issues tracking

Github



Zeppelin is web-based not
You can make beautiful dat

Mahout Tutorial
R Tutorial
Zeppelin Tutorial
Zeppelin Tutorial: Python - matplotlib basic

Create new note

Note Name

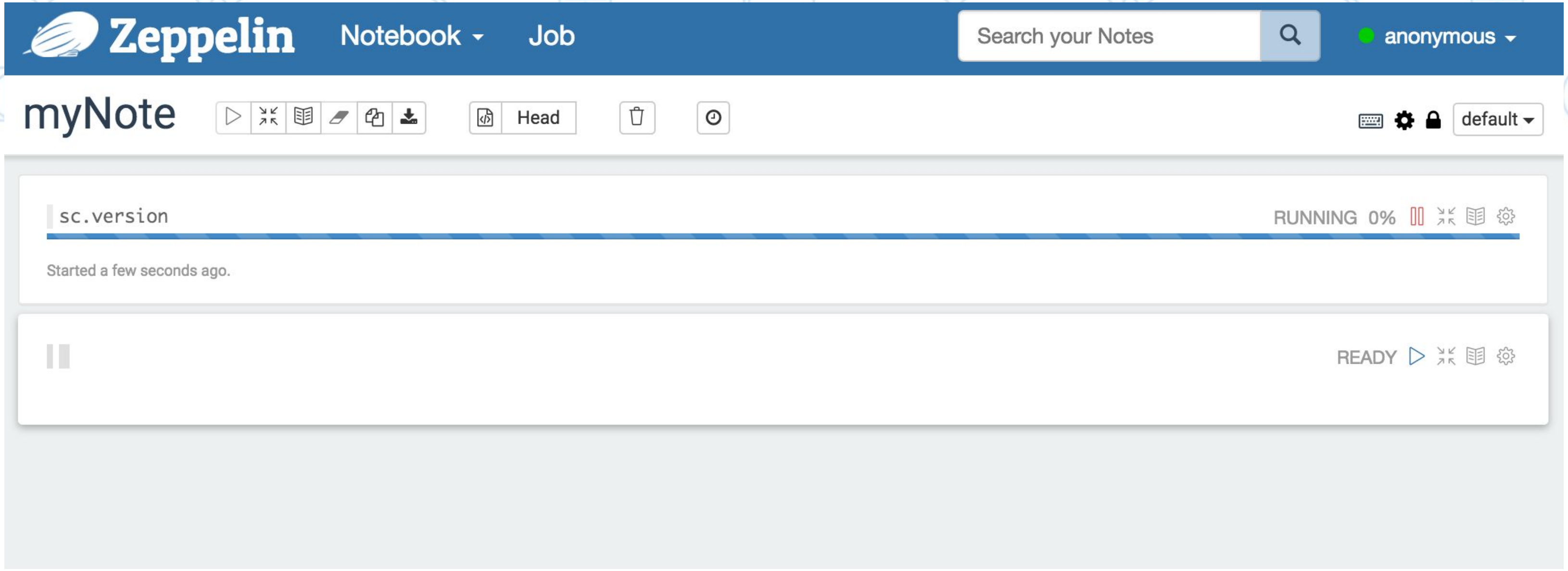
Note name

Default Interpreter --Select--

Use '/' to create folders. Example: /NoteDirA/Notebook1

Create new note

- Select **default** interpreter
- Also can create **directory** with "DirName/NoteName"



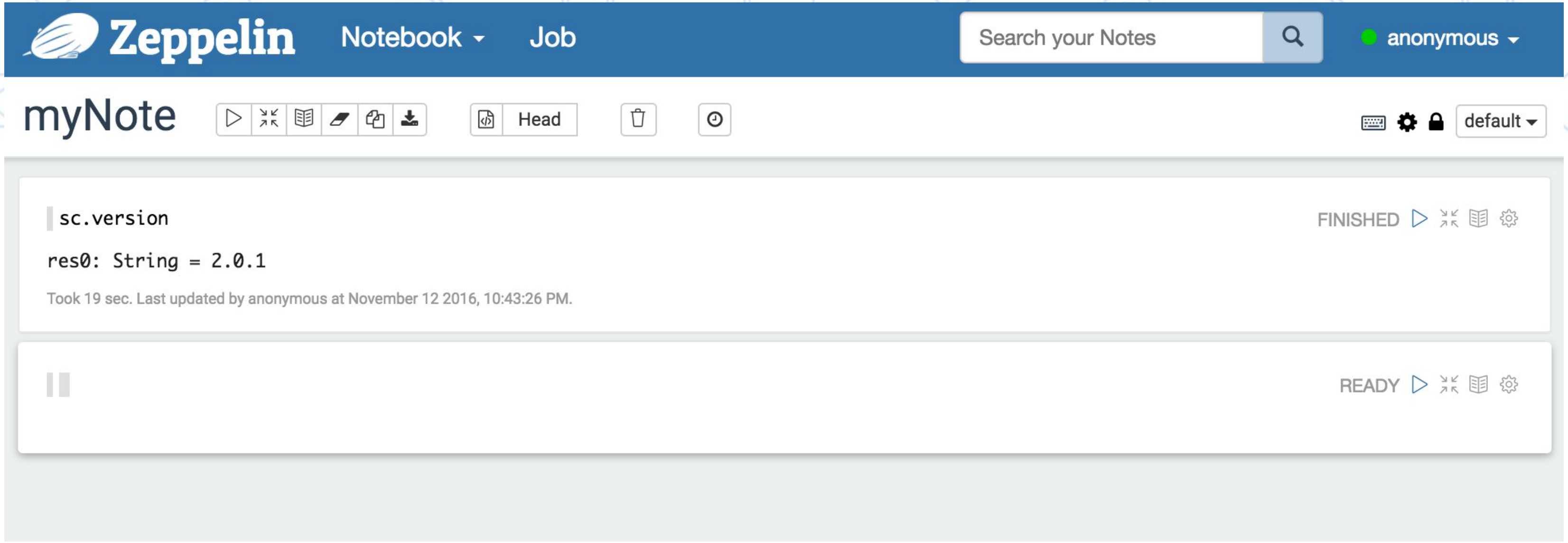
```
sc.version
```

RUNNING 0%

Started a few seconds ago.

Run your code in the paragraph

- Just write down your code and run (shift + enter) when you're using default interpreter
- You can see the progress bar / can stop the paragraph while running

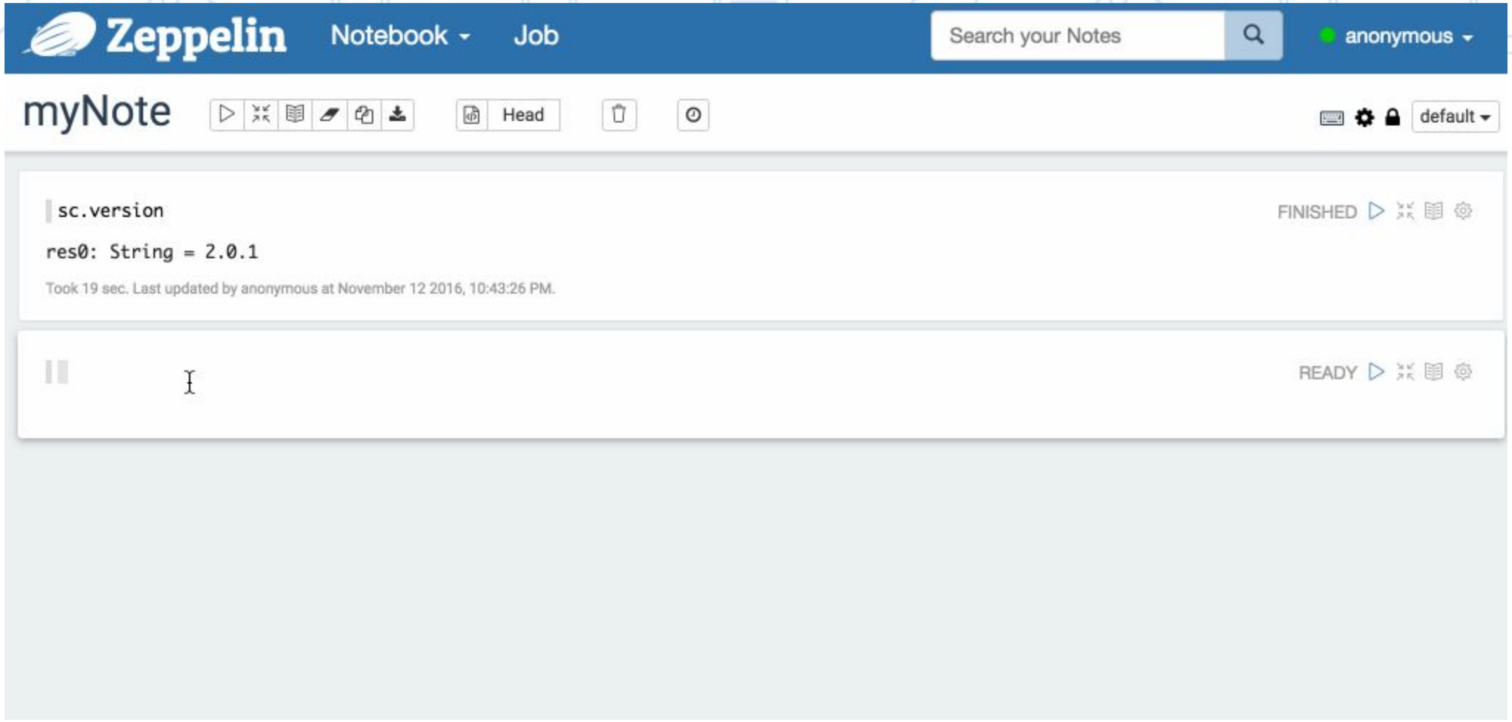


```
sc.version
res0: String = 2.0.1
```

Took 19 sec. Last updated by anonymous at November 12 2016, 10:43:26 PM.

The output after running

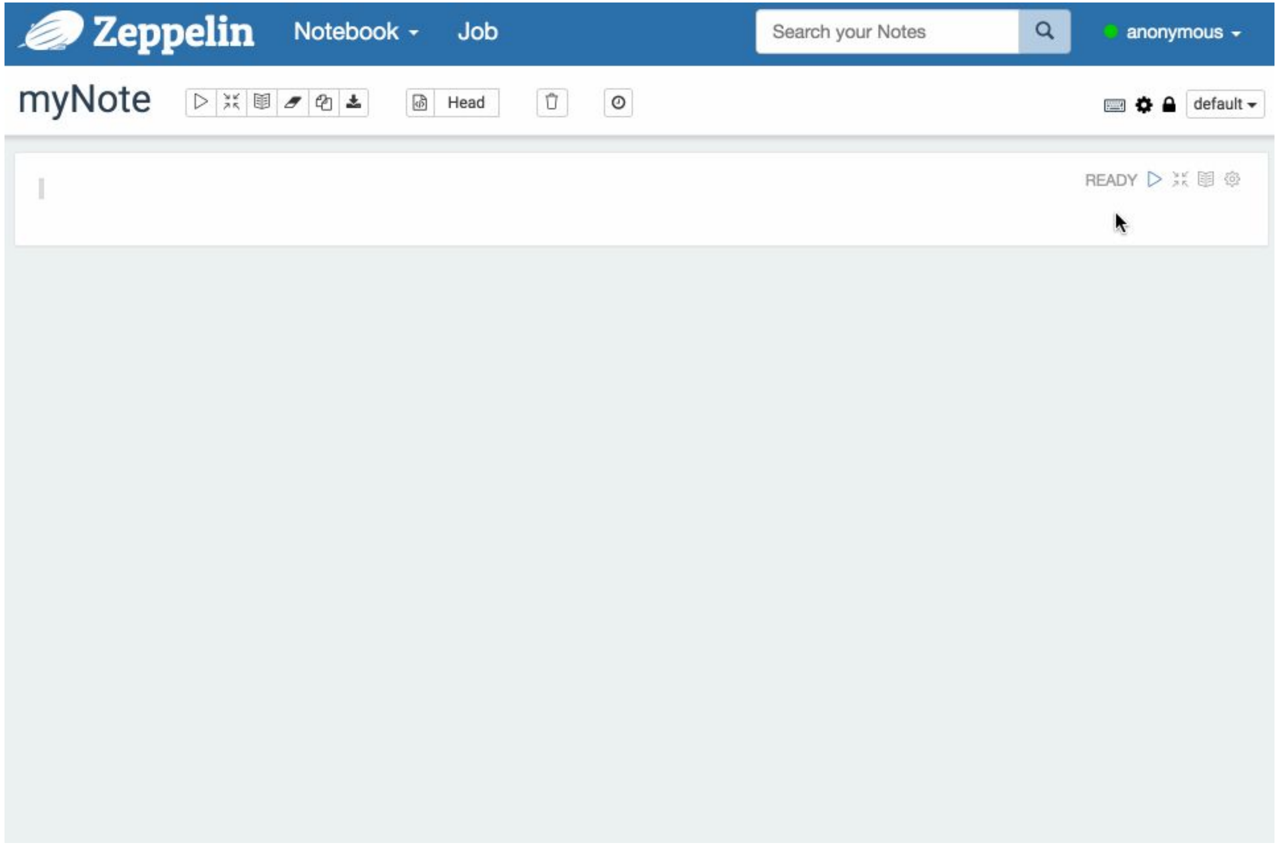
- The result of your code will be printed below the editor with elapsed time/ date/ user name



```
sc.version

res0: String = 2.0.1
```

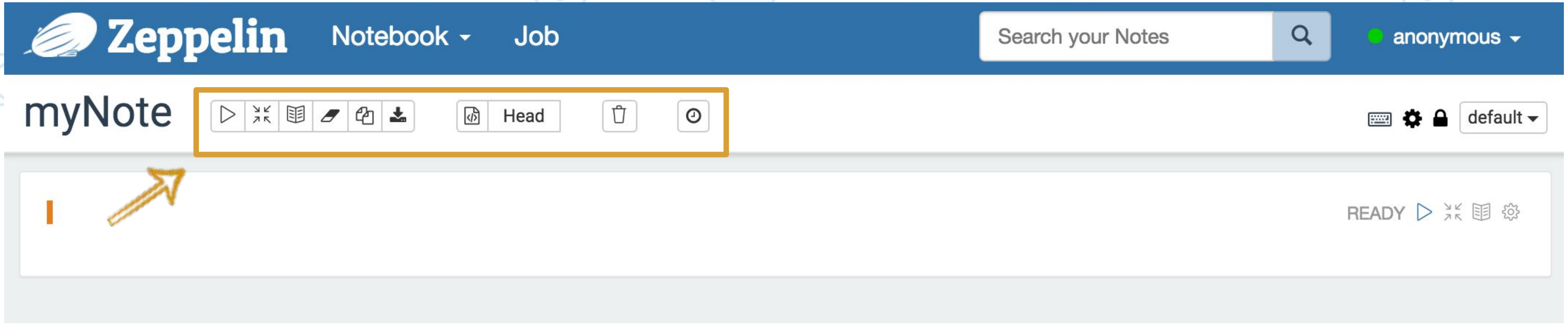
Took 19 sec. Last updated by anonymous at November 12 2016, 10:43:26 PM.

- You can use another interpreter with “%interpreter_name” in the header of paragraph
- E.g. %md for Markdown interpreter to add some description :)



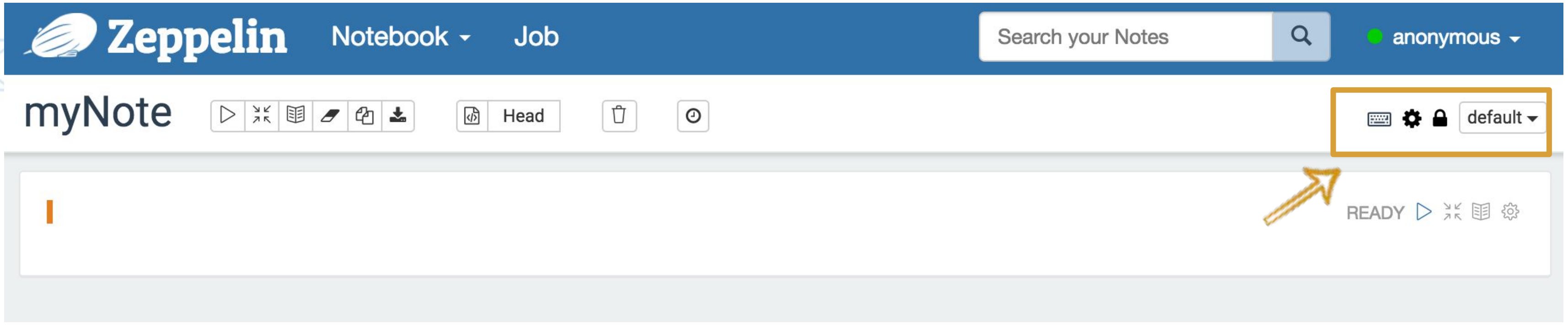
Interpreter Binding

- Change default interpreter
- Activate/deactivate
- Restart



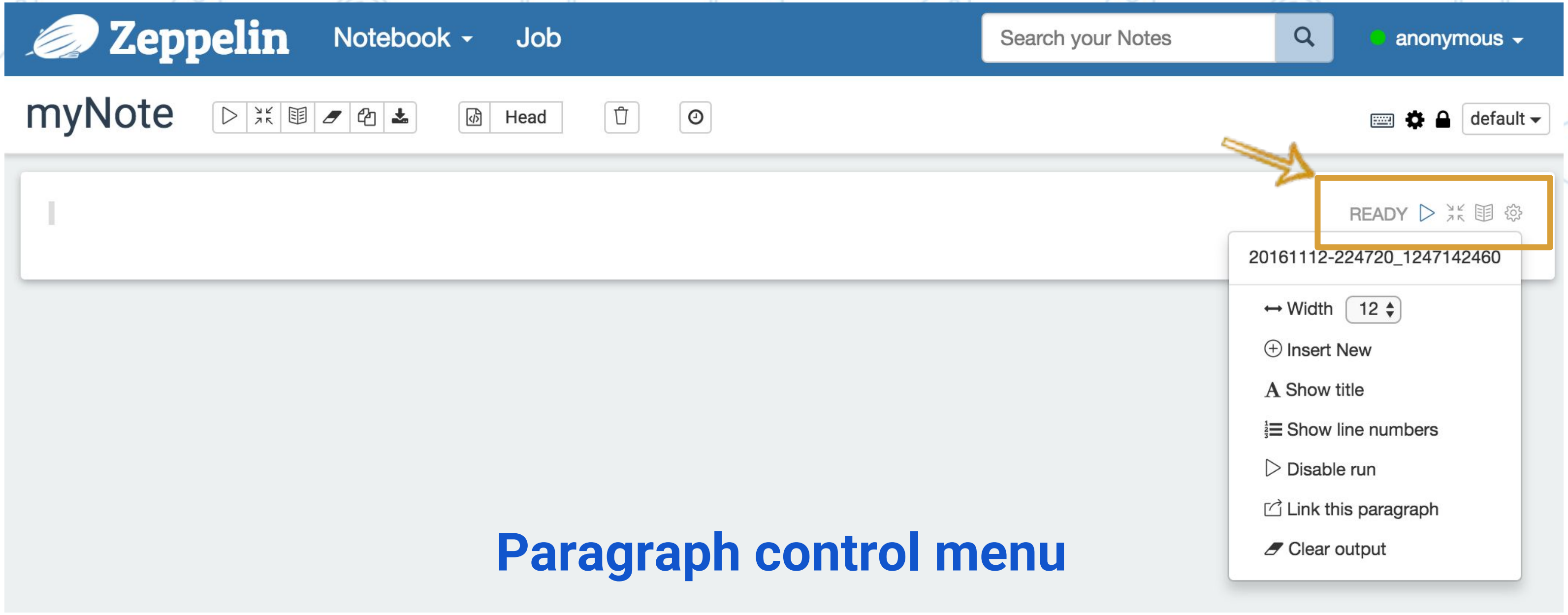
Note control menu

- Run all paragraphs
- Hide or show **code**
- Hide/ show/ clear **the outputs**
- Clone / exports / delete notes
- Note version control
- Run scheduler with **cron expression**



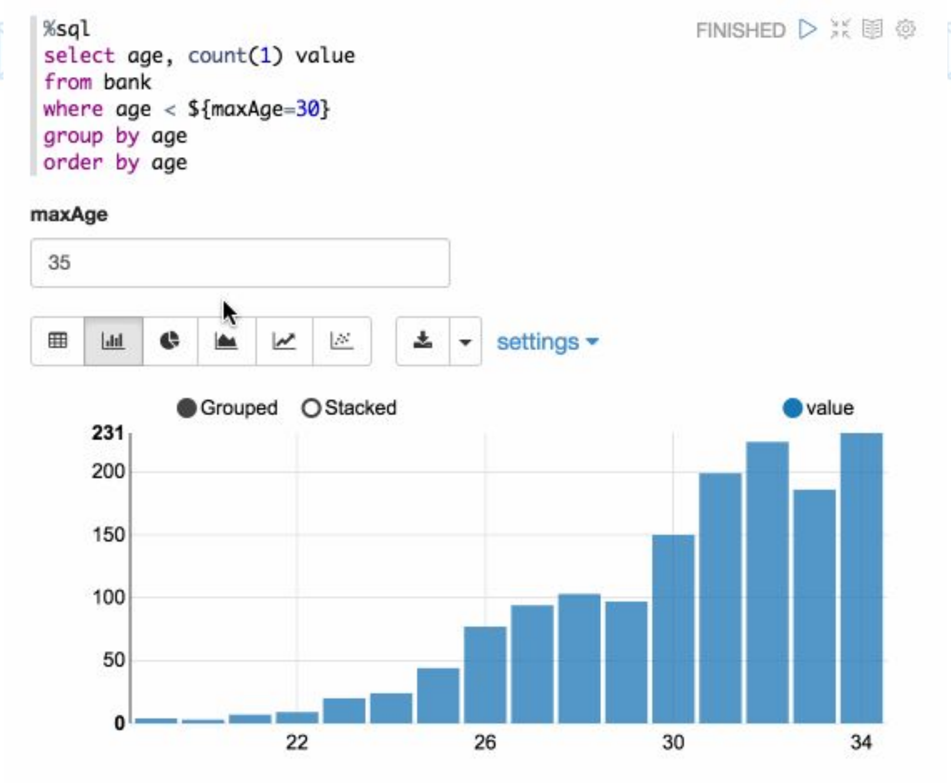
Note control menu

- Keyboard shortcuts
- Interpreter Binding
- Note permission setting
- LookNFeel setting (default / simple / report mode)



Paragraph control menu

- Check the status of result (READY / RUNNING / FINISHED/ ERROR/ PENDING)
- Run ▷ / Stop ▯▯ the paragraph
- Hide or show the code
- Hide or show the output
- Paragraph setting with gear icon

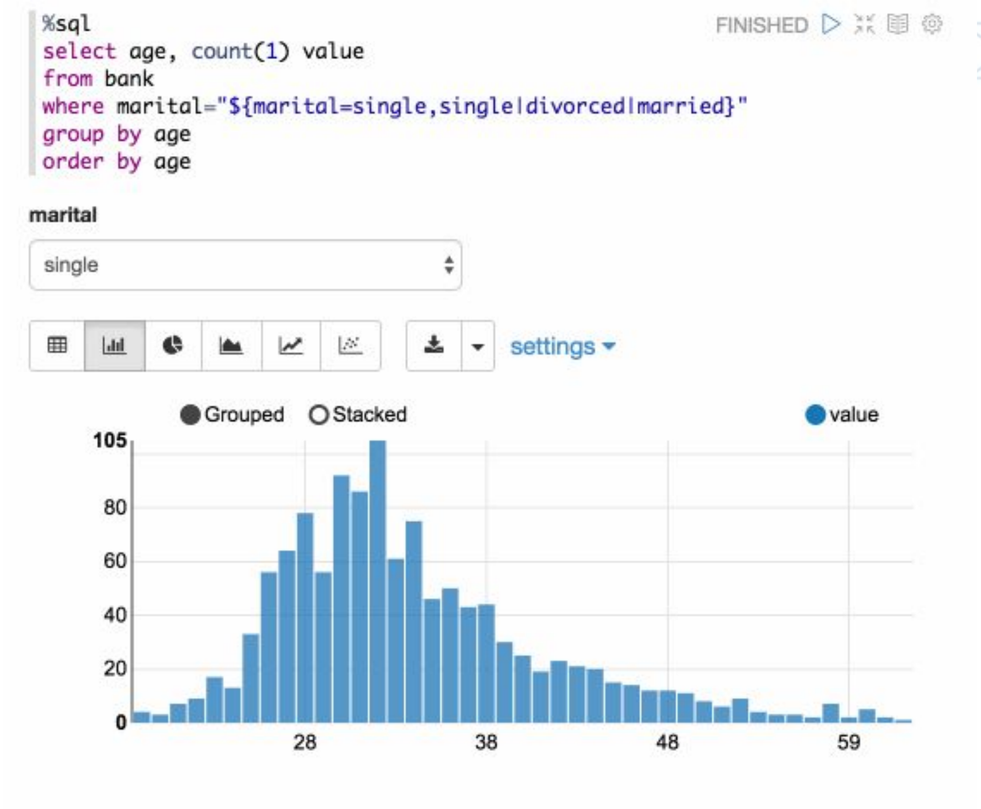


```
%sql
select age, count(1) value
from bank
where age < ${maxAge=30}
group by age
order by age
```

maxAge: 35

The basic feature - #1 Dynamic form

- You can create a dynamic input form in Zeppelin
- Text form: ***${form_name}***
- https://zeppelin.apache.org/docs/latest/manual/dynamicform.html#text-input-form



```
%sql
select age, count(1) value
from bank
where marital="${marital=single,single|divorced|married}"
group by age
order by age
```

The basic feature - #1 Dynamic form

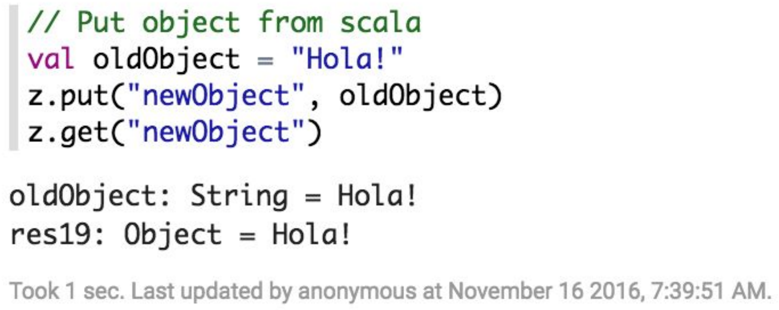
- Select form : ***${form_name = defaultValue, option1|option2...}***
- https://zeppelin.apache.org/docs/latest/manual/dynamicform.html#select-form

The basic feature - #2 Zeppelin Context

- Zeppelin automatically injects ***ZeppelinContext*** as variable "***z***" in your ***Scala*** **&** ***Python*** env.
- https://zeppelin.apache.org/docs/latest/interpreter/spark.html#zeppelincontext

The basic feature - #2 Zeppelin Context

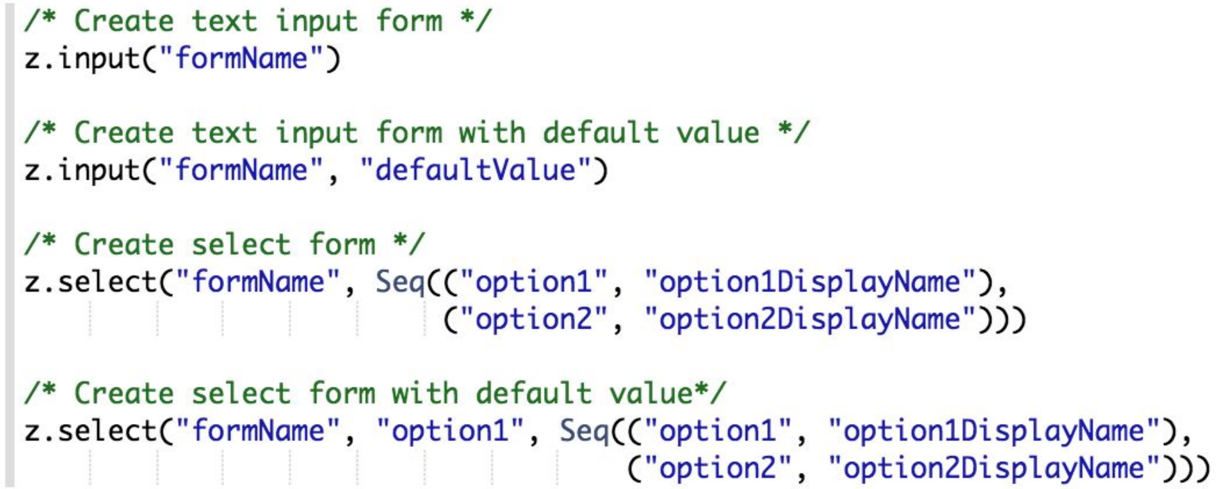
- ***Object Exchange***
- You can put some objects from Scala and read it from Python, vise versa.



```
// Put object from scala
val oldObject = "Hola!"
z.put("newObject", oldObject)
z.get("newObject")
```

```
oldObject: String = Hola!
res19: Object = Hola!
```

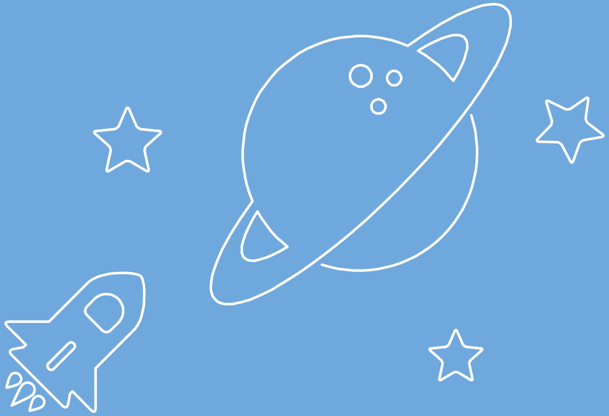
Took 1 sec. Last updated by anonymous at November 16 2016, 7:39:51 AM.



```
/* Create text input form */
z.input("formName")

/* Create text input form with default value */
z.input("formName", "defaultValue")

/* Create select form */
z.select("formName", Seq(("option1", "option1DisplayName"),
                         ("option2", "option2DisplayName")))

/* Create select form with default value*/
z.select("formName", "option1", Seq(("option1", "option1DisplayName"),
                                    ("option2", "option2DisplayName")))
```

The basic feature - #2 Zeppelin Context

- ***Form creation***
- ***ZeppelinContext*** provides function for creating forms.
- You can create forms programmatically in scala/python env.



Hands on Session

Let’s play with example dataset in your Zeppelin notes