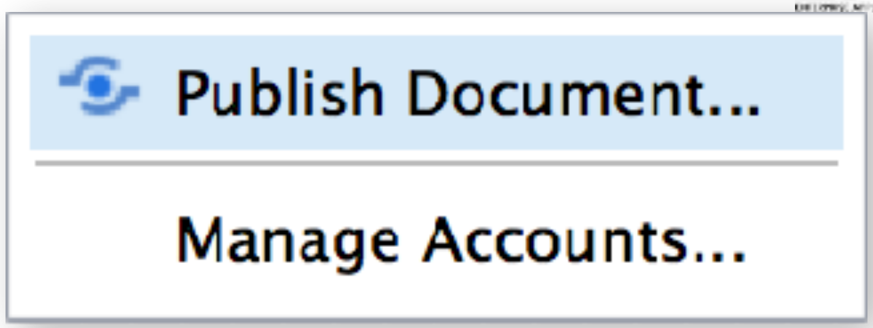
Push-Button Publishing in RStudio Connect

Jeff Allen 11/2017

RStudio

@trestleJeff

RStudio Connect

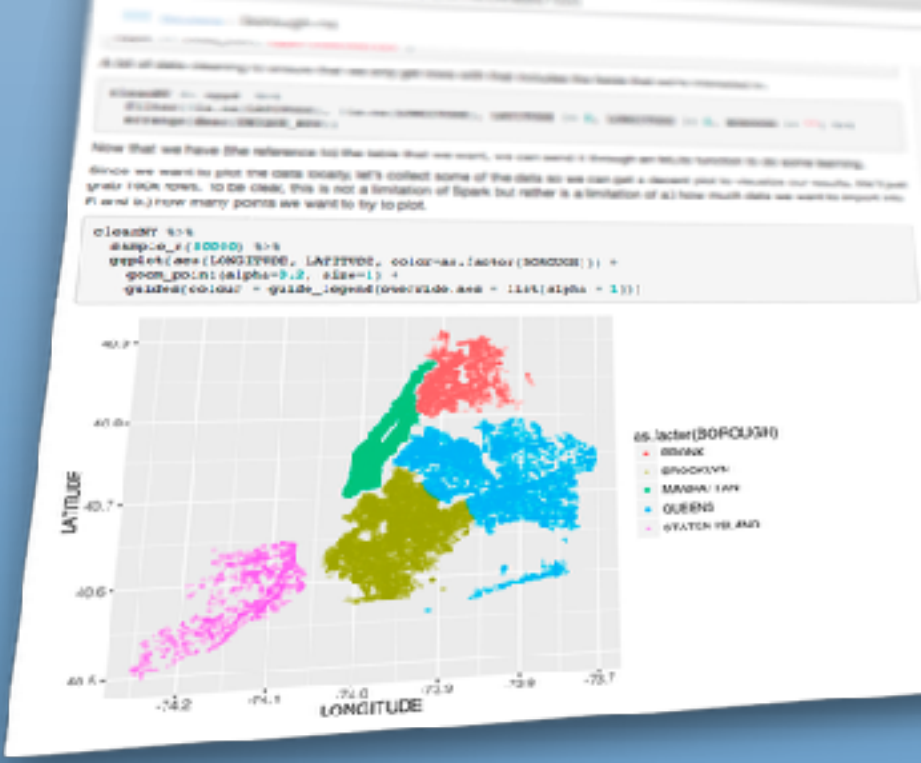


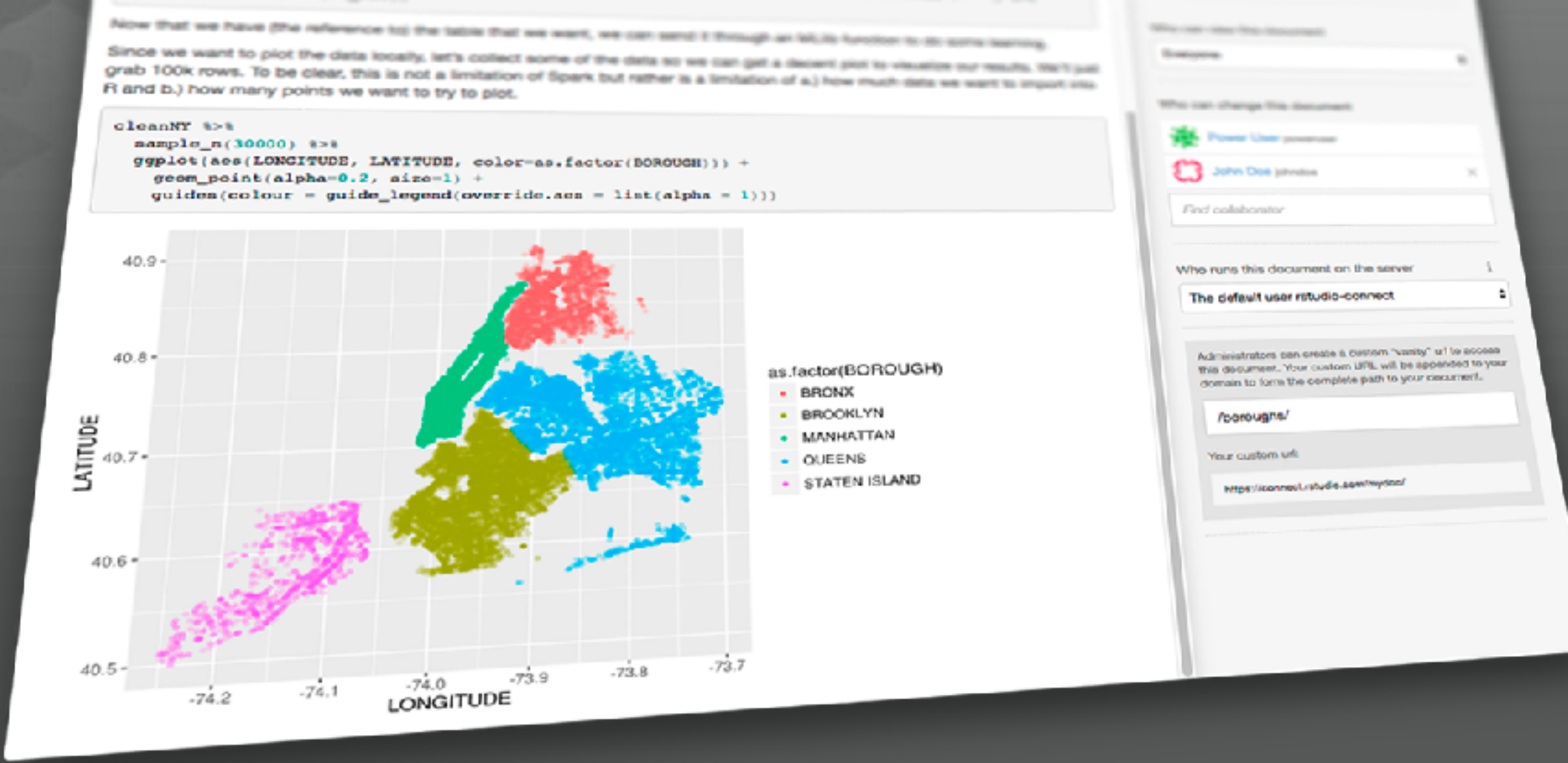
- Push-button publishing from the RStudio IDE
- Manages all the content types you produce in R:

  Shiny, APIs, R Markdown, plots, etc.
- On-premises, commercial
- Share data science artifacts

Three stages of R adoption in an organization

1. Bespoke analysis
2. Interactive tools
3. Fully integrated





Now that we have (the reference to) the table that we want, we can send it through an MLlib function to do some learning.

Since we want to plot the data locally, let's collect some of the data so we can get a decent plot to visualize our results. We'll just grab 100k rows. To be clear, this is not a limitation of Spark but rather is a limitation of a.) how much data we want to import into R and b.) how many points we want to try to plot.

```
cleanNY %>%
  sample_n(30000) %>%
  ggplot(aes(LONGITUDE, LATITUDE, color=as.factor(BOROUGH))) +
    geom_point(alpha=0.2, size=1) +
    guides(colour = guide_legend(override.aes = list(alpha = 1)))
```

Who can view this document

Everyone

Who can change this document

Power User poweruser

John Doe johndoe

Find collaborator

Who runs this document on the server

The default user rstudio-connect

Administrators can create a custom "vanity" url to access this document. Your custom URL will be appended to your domain to form the complete path to your document.

/boroughs/

Your custom url:

https://connect.rstudio.com/mytoc/

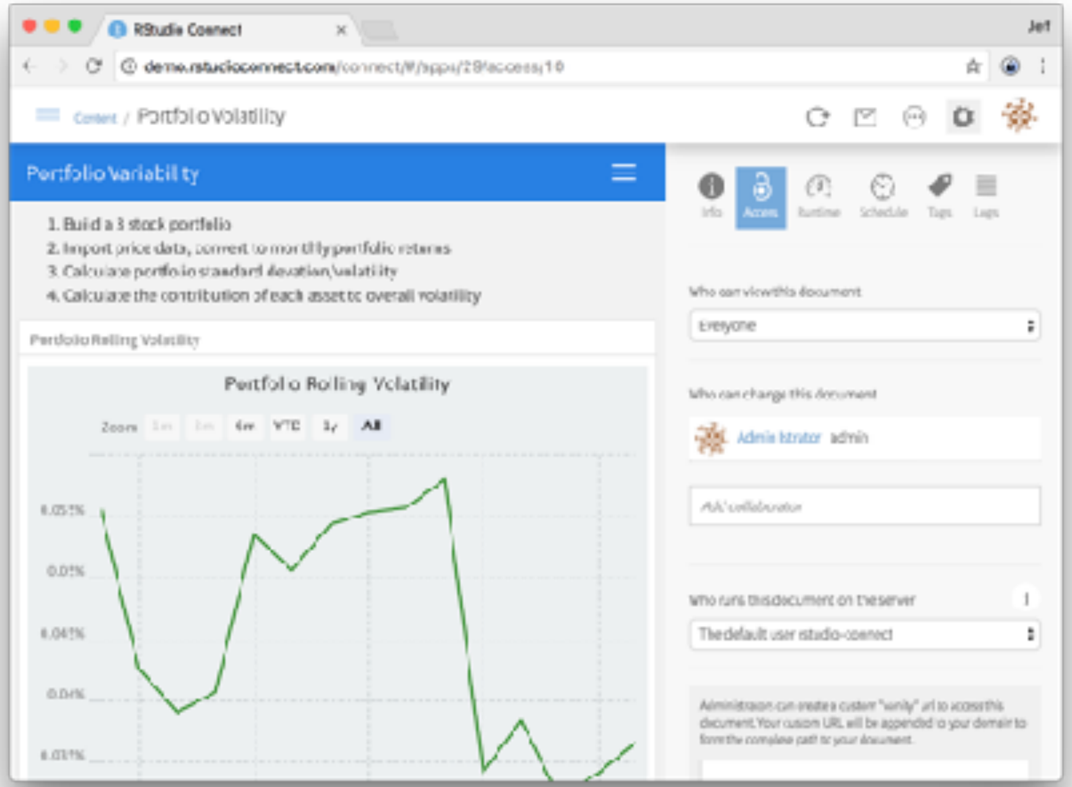
Stage 1: Bespoke Analysis

Stage 1: Overview

- Data scientist is a black box
- Delivers a static result (plot, table, document)
- Revise question or ask new questions and start over
- Individual secretly installing R on your laptop to a group managing their own RStudio Server

Stage 1: Use Cases

- Sales reporting
- Investment analysis, reproducible decision making
- Game usage metrics reporting



demo.rstudioconnect.com/connect/#/apps/29/access/10

1. Build a 3 stock portfolio
2. Import price data, convert to monthly portfolio returns
3. Calculate portfolio standard deviation/volatility
4. Calculate the contribution of each asset to overall volatility

Portfolio Rolling Volatility

Portfolio Rolling Volatility

Who can view this document: Everyone

Who can change this document: Administrator admin

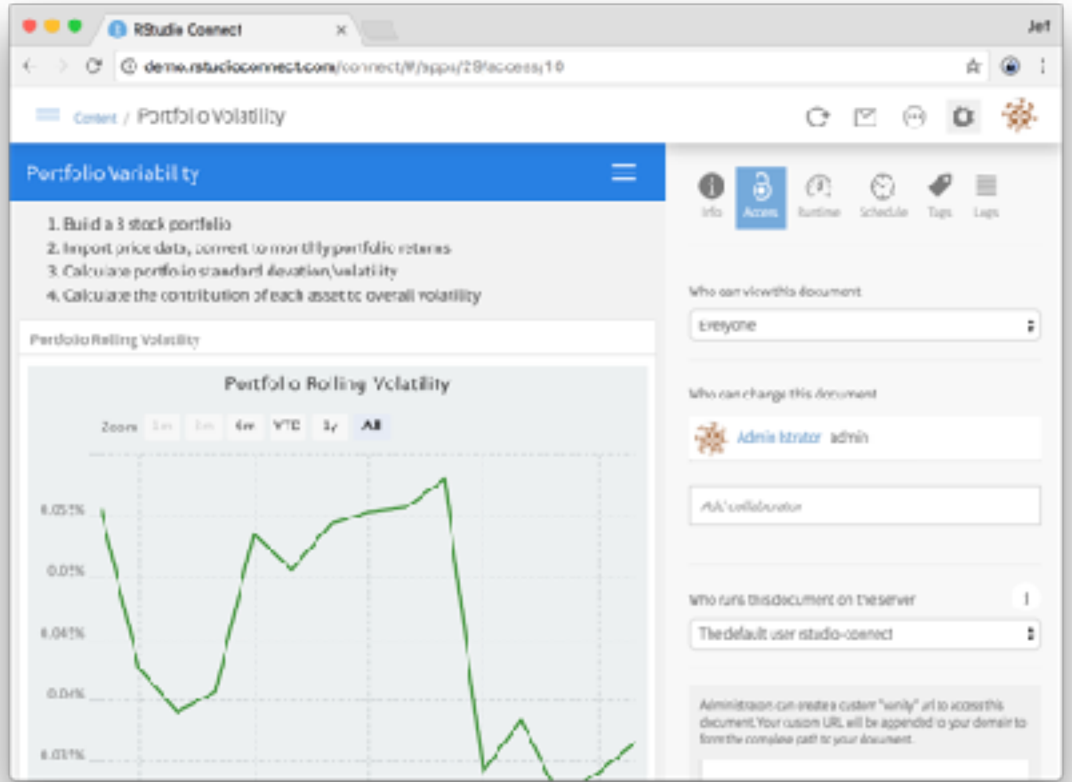
Add collaborator

Who runs this document on the server: The default user rstudio-connect

Administrators can create a custom "vanity" url to access this document. Your custom URL will be appended to your domain to form the complete path to your document.

Stage 1: Issues

- Disorganized output
- Managing access
- Irreproducible
- No ability to automate


Portfolio Variability

1. Build a 3 stock portfolio
2. Import price data, convert to monthly portfolio returns
3. Calculate portfolio standard deviation/volatility
4. Calculate the contribution of each asset to overall volatility

Who can view this document: Everyone

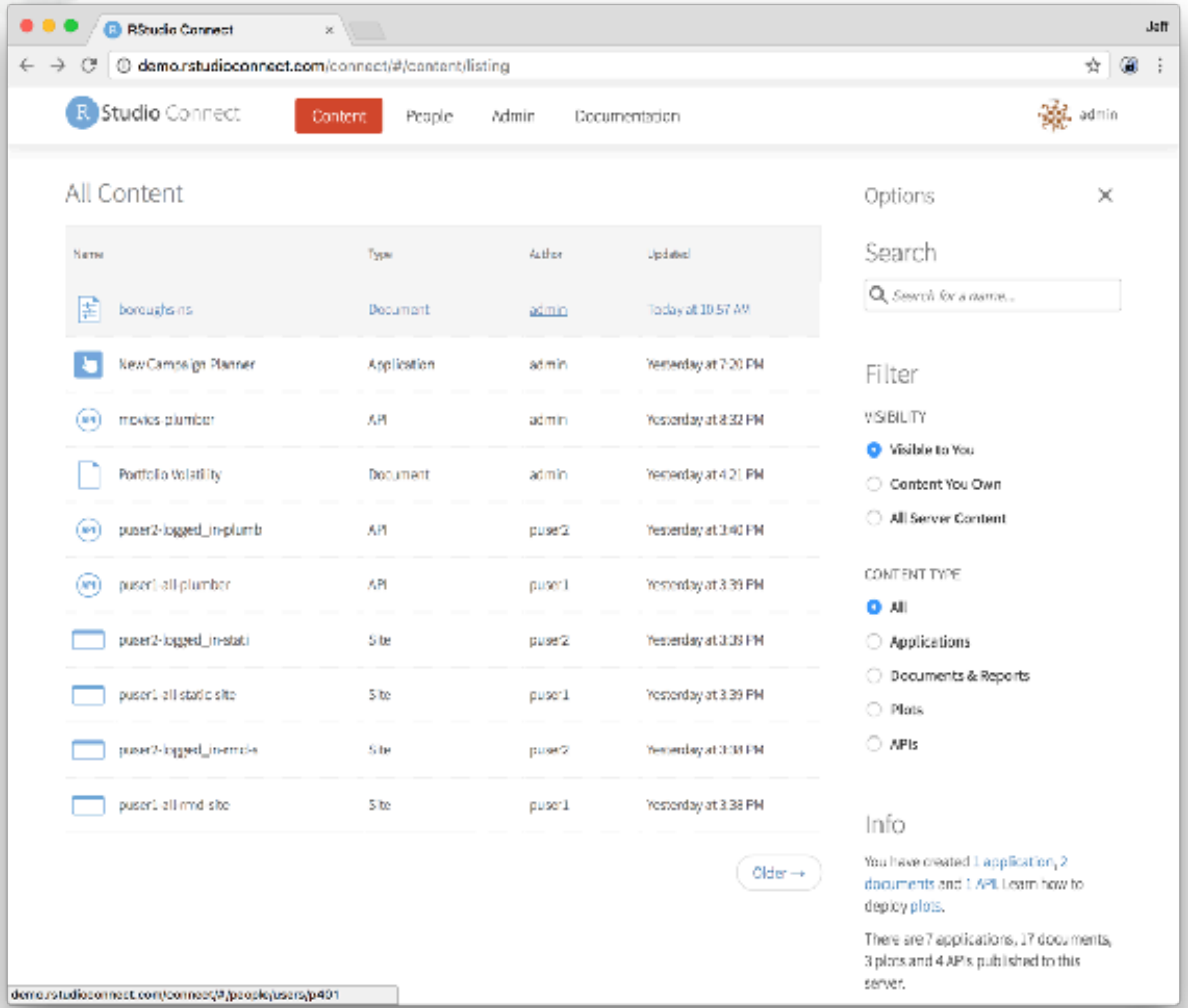
Who can change this document: Administrator admin

Who runs this document on the server: The default user rstudio-connect

Administrators can create a custom "vanity" url to access this document. Your custom URL will be appended to your domain to form the complete path to your document.

RStudio Connect
demo.rstudioconnect.com/connect/#/apps/28/access/10
Content / Portfolio Volatility
Info
Access
Runtime
Schedule
Tags
Logs
Portfolio Rolling Volatility
Zoom 1m 3m 6m YTD 1y All
Add collaborator

Stage 1 Demo



All Content

Name	Type	Author	Updated
boroughs.rns	Document	admin	Today at 10:57 AM
New Campaign Planner	Application	admin	Yesterday at 7:20 PM
movies-plumber	API	admin	Yesterday at 8:32 PM
Portfolio Volatility	Document	admin	Yesterday at 4:21 PM
puser2-logged_in-plumb	API	puser2	Yesterday at 3:40 PM
puser1-all-plumber	API	puser1	Yesterday at 3:39 PM
puser2-logged_in-stati	Site	puser2	Yesterday at 3:39 PM
puser1-all-static-site	Site	puser1	Yesterday at 3:39 PM
puser2-logged_in-rmd-s	Site	puser2	Yesterday at 3:38 PM
puser1-all-rmd-site	Site	puser1	Yesterday at 3:38 PM

Older →

Options

Search

Search for a name...

Filter

VISIBILITY

[✓ Visible to You]
Content You Own
All Server Content

CONTENT TYPE

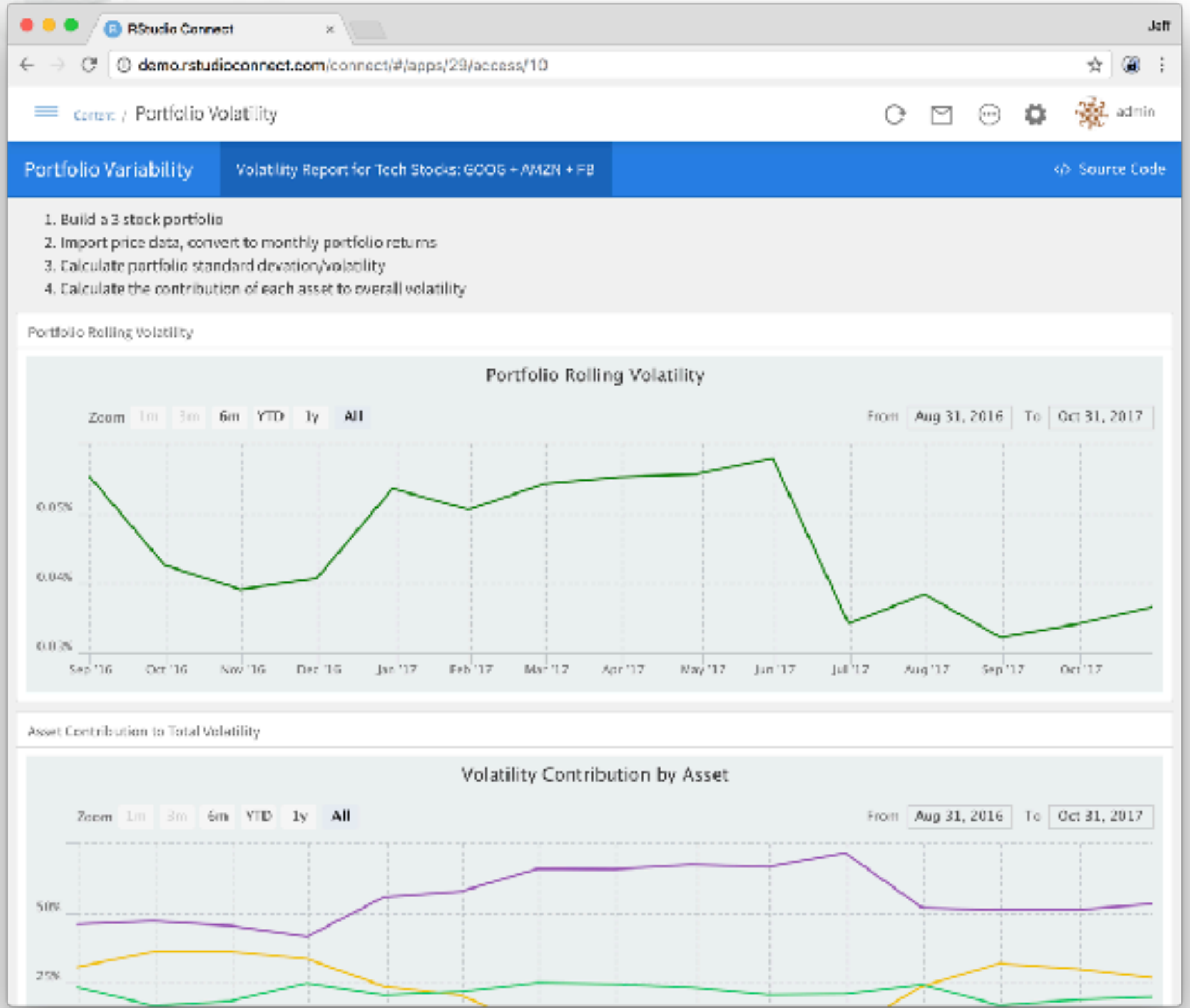
[✓ All]
Applications
Documents & Reports
Plots
APIs

Info

You have created 1 application, 2 documents and 1 API. Learn how to deploy plots.

There are 7 applications, 17 documents, 3 plots and 4 APIs published to this server.

demo.rstudioconnect.com/connect/#/people/users/p401


RStudio Connect

demo.rstudioconnect.com/connect/#/apps/29/access/10

Content / Portfolio Volatility

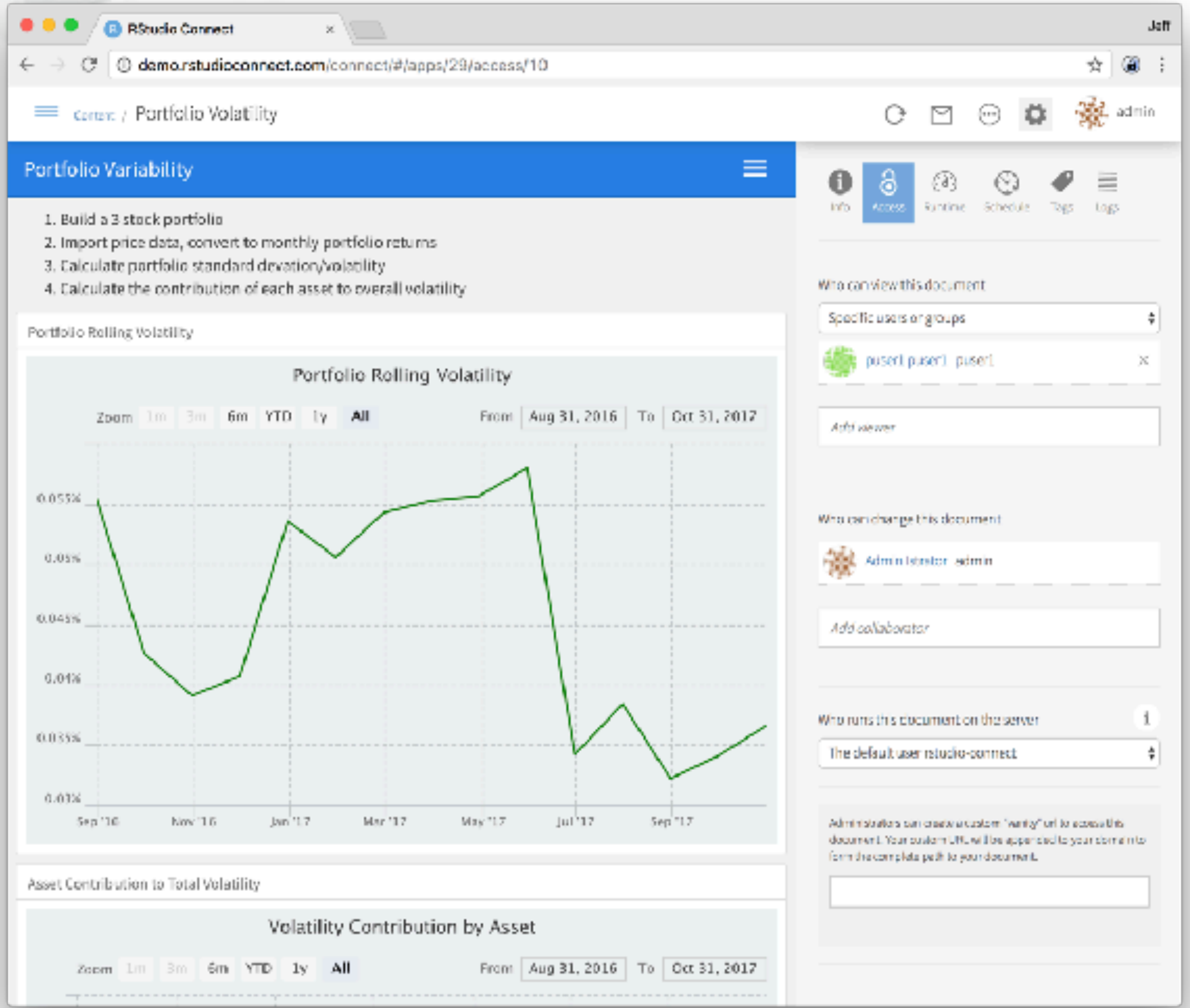
Portfolio Variability | Volatility Report for Tech Stocks: GOOG + AMZN + FB | Source Code

1. Build a 3 stock portfolio
2. Import price data, convert to monthly portfolio returns
3. Calculate portfolio standard deviation/volatility
4. Calculate the contribution of each asset to overall volatility

Portfolio Rolling Volatility

Asset Contribution to Total Volatility

Volatility Contribution by Asset
Zoom 1m 3m 6m YTD 1y All
From Aug 31, 2016 To Oct 31, 2017
50%
25%


RStudio Connect
demo.rstudioconnect.com/connect/#/apps/29/access/10
Content / Portfolio Volatility
admin
Info
Access
Runtime
Schedule
Tags
Logs

Portfolio Variability

1. Build a 3 stock portfolio
2. Import price data, convert to monthly portfolio returns
3. Calculate portfolio standard deviation/volatility
4. Calculate the contribution of each asset to overall volatility

Portfolio Rolling Volatility

Portfolio Rolling Volatility
Zoom 1m 3m 6m YTD 1y All
From Aug 31, 2016 To Oct 31, 2017
0.055%
0.05%
0.045%
0.04%
0.035%
0.03%
Sep '16
Nov '16
Jan '17
Mar '17
May '17
Jul '17
Sep '17

Asset Contribution to Total Volatility

Volatility Contribution by Asset
Zoom 1m 3m 6m YTD 1y All
From Aug 31, 2016 To Oct 31, 2017

Who can view this document

Specific users or groups

puser1 puser1 puser1

Add viewer

Who can change this document

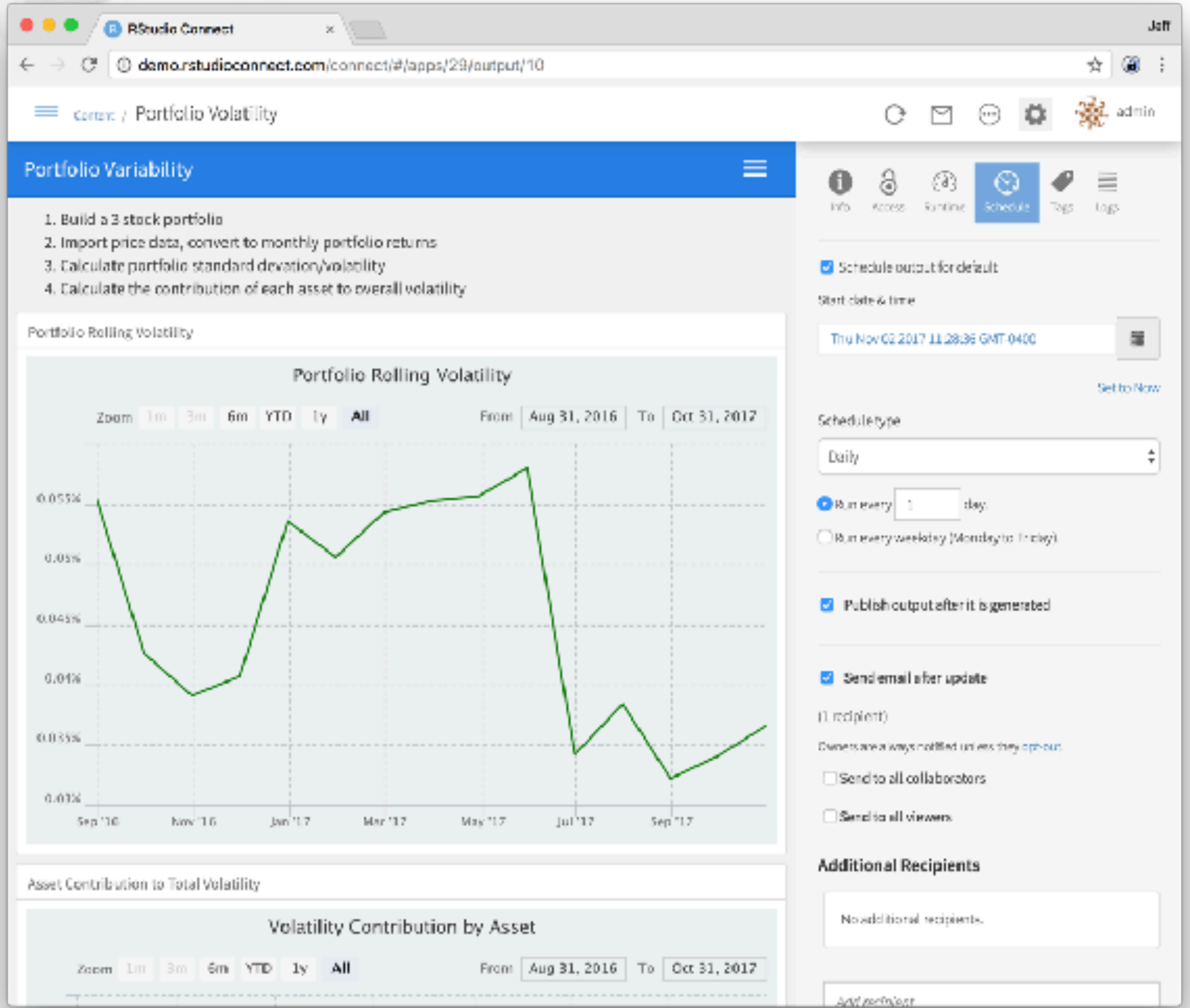
Administrator admin

Add collaborator

Who runs this document on the server

The default user rstudio-connect

Administrators can create a custom "vanity" url to access this document. Your custom URL will be appended to your domain to form the complete path to your document.



Portfolio Variability

1. Build a 3 stock portfolio
2. Import price data, convert to monthly portfolio returns
3. Calculate portfolio standard deviation/volatility
4. Calculate the contribution of each asset to overall volatility

Portfolio Rolling Volatility

Portfolio Rolling Volatility

Asset Contribution to Total Volatility

Volatility Contribution by Asset

Zoom 1m 3m 6m YTD 1y All From Aug 31, 2016 To Oct 31, 2017

- [✓] Schedule output for default

Start date & time: Thu Nov 02 2017 11:28:36 GMT-0400

Set to Now

Schedule type: Daily

- [✓ Run every 1 day]
- Run every weekday (Monday to Friday)

- [✓] Publish output after it is generated

- [✓] Send email after update

(1 recipient)

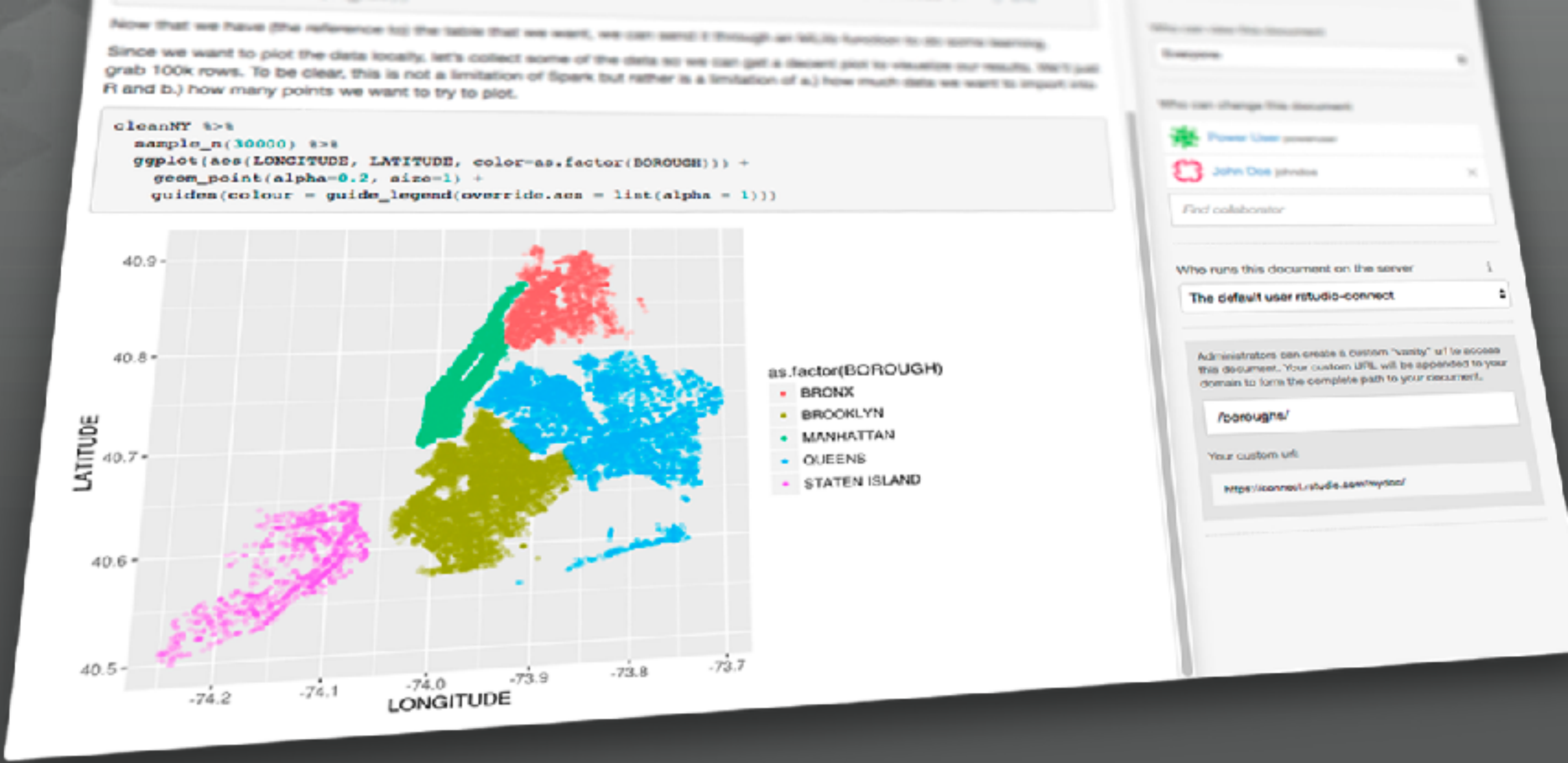
Owners are always notified unless they opt-out.

- [ ] Send to all collaborators
- [ ] Send to all viewers

Additional Recipients

No additional recipients.

Add recipient



Now that we have (the reference to) the table that we want, we can send it through an MLlib function to do some learning.

Since we want to plot the data locally, let's collect some of the data so we can get a decent plot to visualize our results. We'll just grab 100k rows. To be clear, this is not a limitation of Spark but rather is a limitation of a.) how much data we want to import into R and b.) how many points we want to try to plot.

```
cleanNY %>%
  sample_n(30000) %>%
  ggplot(aes(LONGITUDE, LATITUDE, color=as.factor(BOROUGH))) +
    geom_point(alpha=0.2, size=1) +
    guides(colour = guide_legend(override.aes = list(alpha = 1)))
```

Who can view this document

Everyone

Who can change this document

Power User poweruser

John Doe johndoe

Find collaborator

Who runs this document on the server

The default user rstudio-connect

Administrators can create a custom "vanity" url to access this document. Your custom URL will be appended to your domain to form the complete path to your document.

/boroughs/

Your custom url:

https://connect.rstudio.com/mydoc/

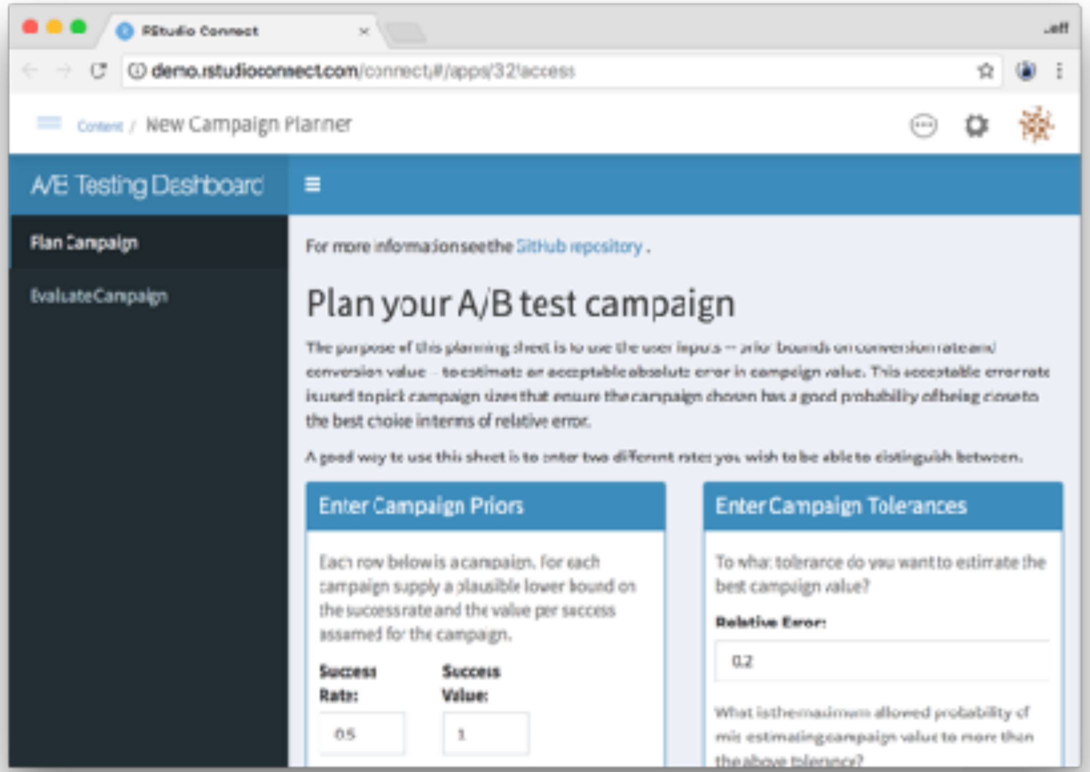
Stage 2: Interactive Tools

Stage 2: Overview

- Some formal acceptance of R
  - Internal user-groups, training
  - Optimizing the on-boarding experience
- Non-R-users consume tools written in R
  - Self-service
  - Shiny, Parameterized R Markdown

Stage 2: Use Cases

- BI dashboarding
- Monitoring & analysis of customer support
- Domain-specific tooling
  - Split testing


RStudio Connect

demo.rstudioconnect.com/connect/#/apps/32/access

Content / New Campaign Planner

A/B Testing Dashboard

Plan Campaign

Evaluate Campaign

For more information see the GitHub repository.

Plan your A/B test campaign

The purpose of this planning sheet is to use the user inputs -- prior bounds on conversion rate and conversion value -- to estimate an acceptable absolute error in campaign value. This acceptable error rate is used to pick campaign sizes that ensure the campaign chosen has a good probability of being close to the best choice in terms of relative error.

A good way to use this sheet is to enter two different rates you wish to be able to distinguish between.

Enter Campaign Priors

Each row below is a campaign. For each campaign supply a plausible lower bound on the success rate and the value per success assumed for the campaign.

Success Rate: 0.5

Success Value: 1

Enter Campaign Tolerances

To what tolerance do you want to estimate the best campaign value?

Relative Error: 0.2

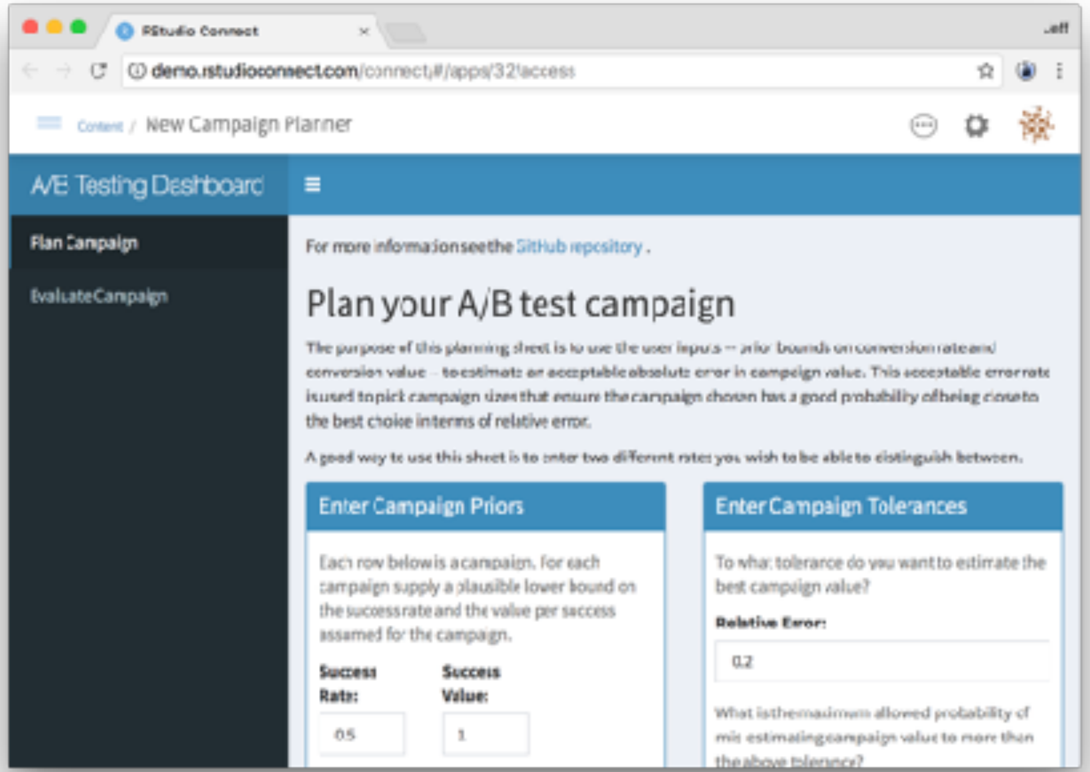
What is the maximum allowed probability of mis-estimating campaign value to more than the above tolerance?

https://github.com/WinVector/CampaignPlanner_v3

Stage 2: Issues

- Hosting
  - Internal vs external
  - Authentication
  - Scalability & performance
- Publishing applications
- Testing

Stage 2 Demo



RStudio Connect

demo.rstudioconnect.com/connect/#/apps/32/access

Content / New Campaign Planner

A/B Testing Dashboard

Plan Campaign

Evaluate Campaign

For more information see the GitHub repository .

Plan your A/B test campaign

The purpose of this planning sheet is to use the user inputs -- prior bounds on conversion rate and conversion value -- to estimate an acceptable absolute error in campaign value. This acceptable error rate is used to pick campaign sizes that ensure the campaign chosen has a good probability of being close to the best choice in terms of relative error.

A good way to use this sheet is to enter two different rates you wish to be able to distinguish between.

Enter Campaign Priors

Each row below is a campaign. For each campaign supply a plausible lower bound on the success rate and the value per success assumed for the campaign.

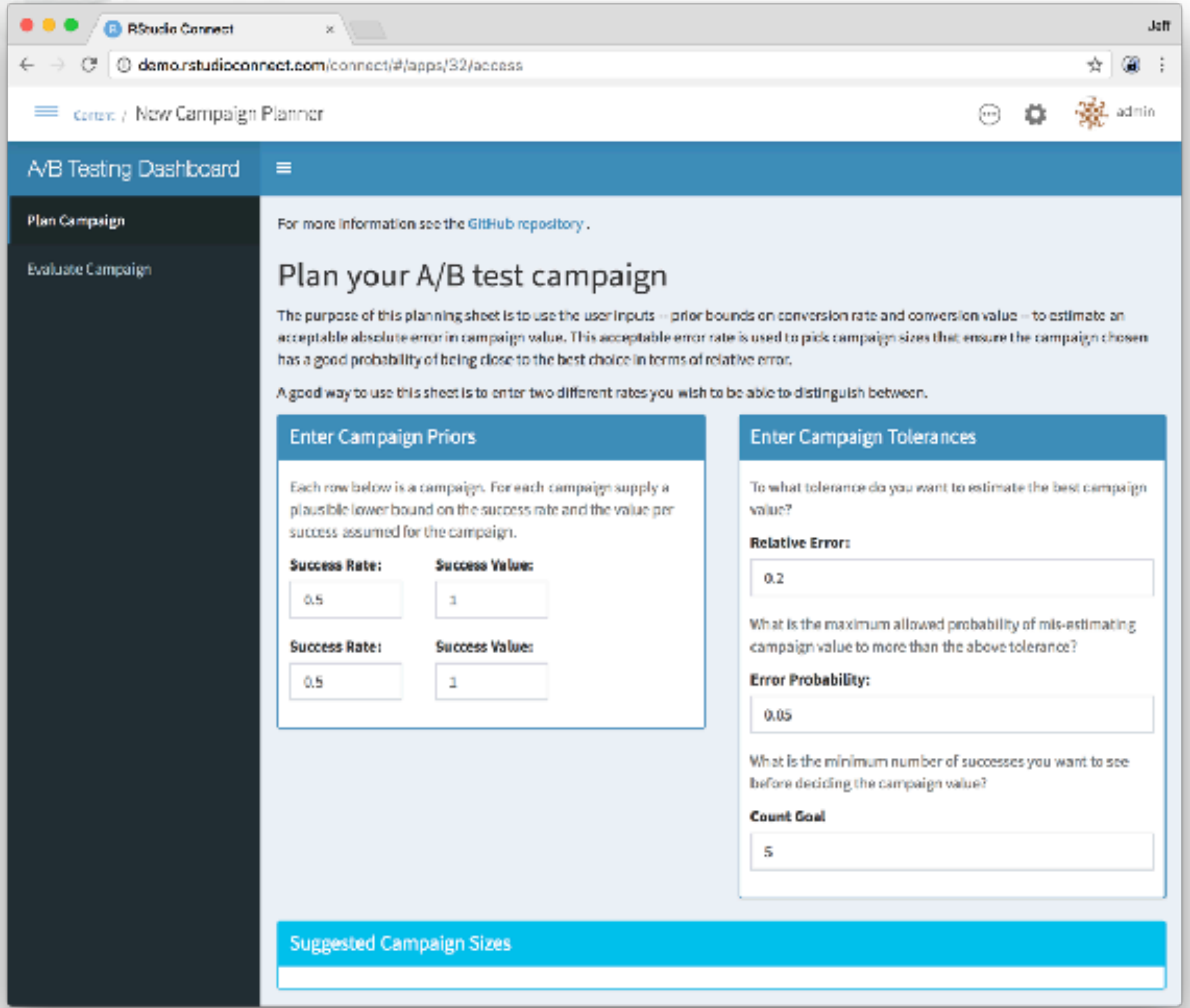
Success Rate:	Success Value:
0.5	1

Enter Campaign Tolerances

To what tolerance do you want to estimate the best campaign value?

Relative Error: 0.2

What is the maximum allowed probability of mis-estimating campaign value to more than the above tolerance?


RStudio Connect
Jeff
demo.rstudioconnect.com/connect/#/apps/32/access
Content / New Campaign Planner
admin
A/B Testing Dashboard
Plan Campaign
Evaluate Campaign

For more information see the GitHub repository .

Plan your A/B test campaign

The purpose of this planning sheet is to use the user inputs -- prior bounds on conversion rate and conversion value -- to estimate an acceptable absolute error in campaign value. This acceptable error rate is used to pick campaign sizes that ensure the campaign chosen has a good probability of being close to the best choice in terms of relative error.

A good way to use this sheet is to enter two different rates you wish to be able to distinguish between.

Enter Campaign Priors

Each row below is a campaign. For each campaign supply a plausible lower bound on the success rate and the value per success assumed for the campaign.

Success Rate:	Success Value:
0.5	1
0.5	1

Enter Campaign Tolerances

To what tolerance do you want to estimate the best campaign value?

Relative Error: 0.2

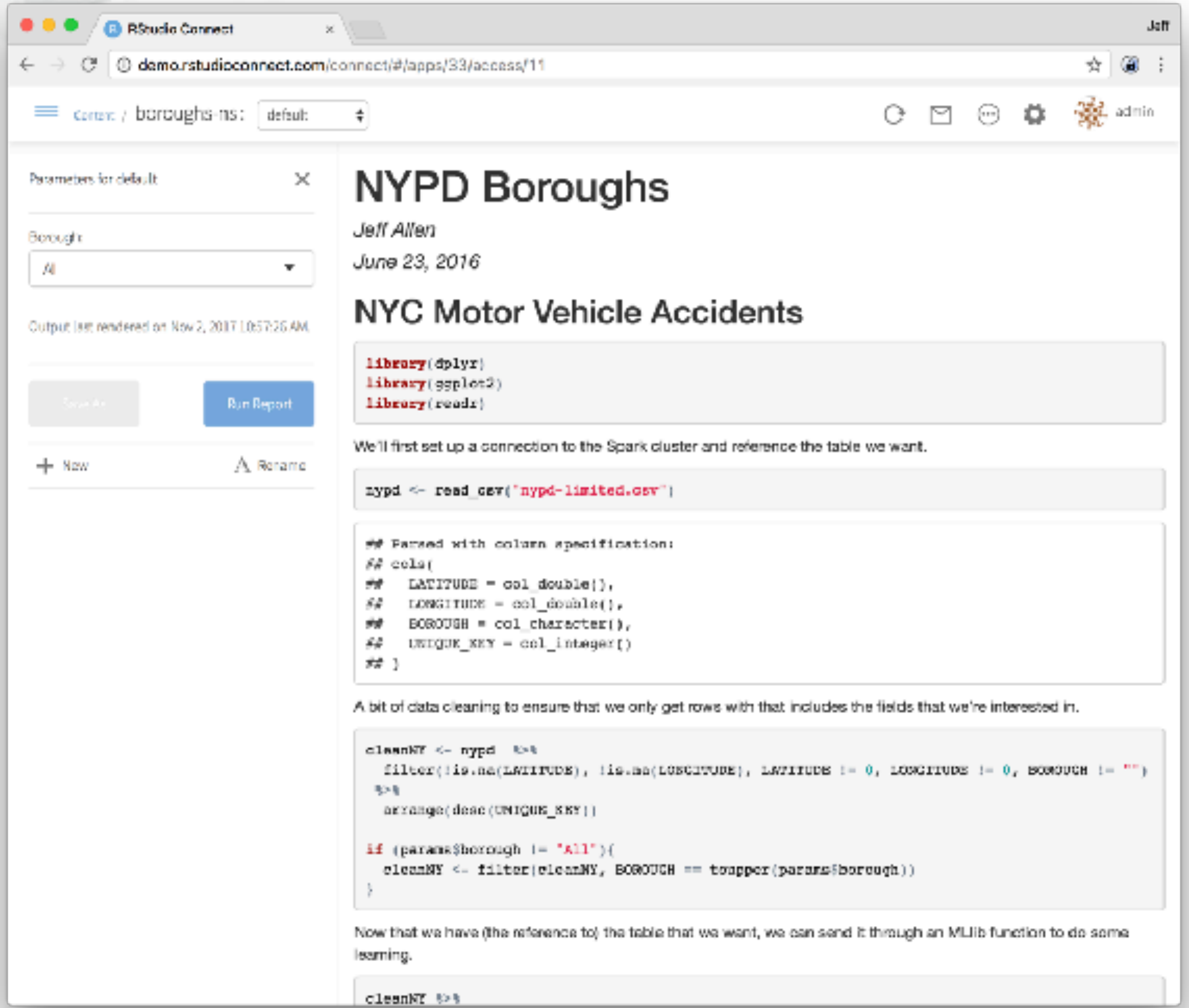
What is the maximum allowed probability of mis-estimating campaign value to more than the above tolerance?

Error Probability: 0.05

What is the minimum number of successes you want to see before deciding the campaign value?

Count Goal 5

Suggested Campaign Sizes


RStudio Connect
Jeff
demo.rstudioconnect.com/connect/#/apps/33/access/11
Content / boroughs-ns : default
admin
Parameters for default
Borough
All
Output last rendered on Nov 2, 2017 10:57:26 AM
Save As
Run Report
New
Rename

NYPD Boroughs

Jeff Allen

June 23, 2016

NYC Motor Vehicle Accidents

```
library(dplyr)
library(ggplot2)
library(readr)
```

We'll first set up a connection to the Spark cluster and reference the table we want.

```
nypd <- read_csv("nypd-limited.csv")
```

```
## Parsed with column specification:
## cols(
##   LATITUDE = col_double(),
##   LONGITUDE = col_double(),
##   BOROUGH = col_character(),
##   UNIQUE_KEY = col_integer()
## )
```

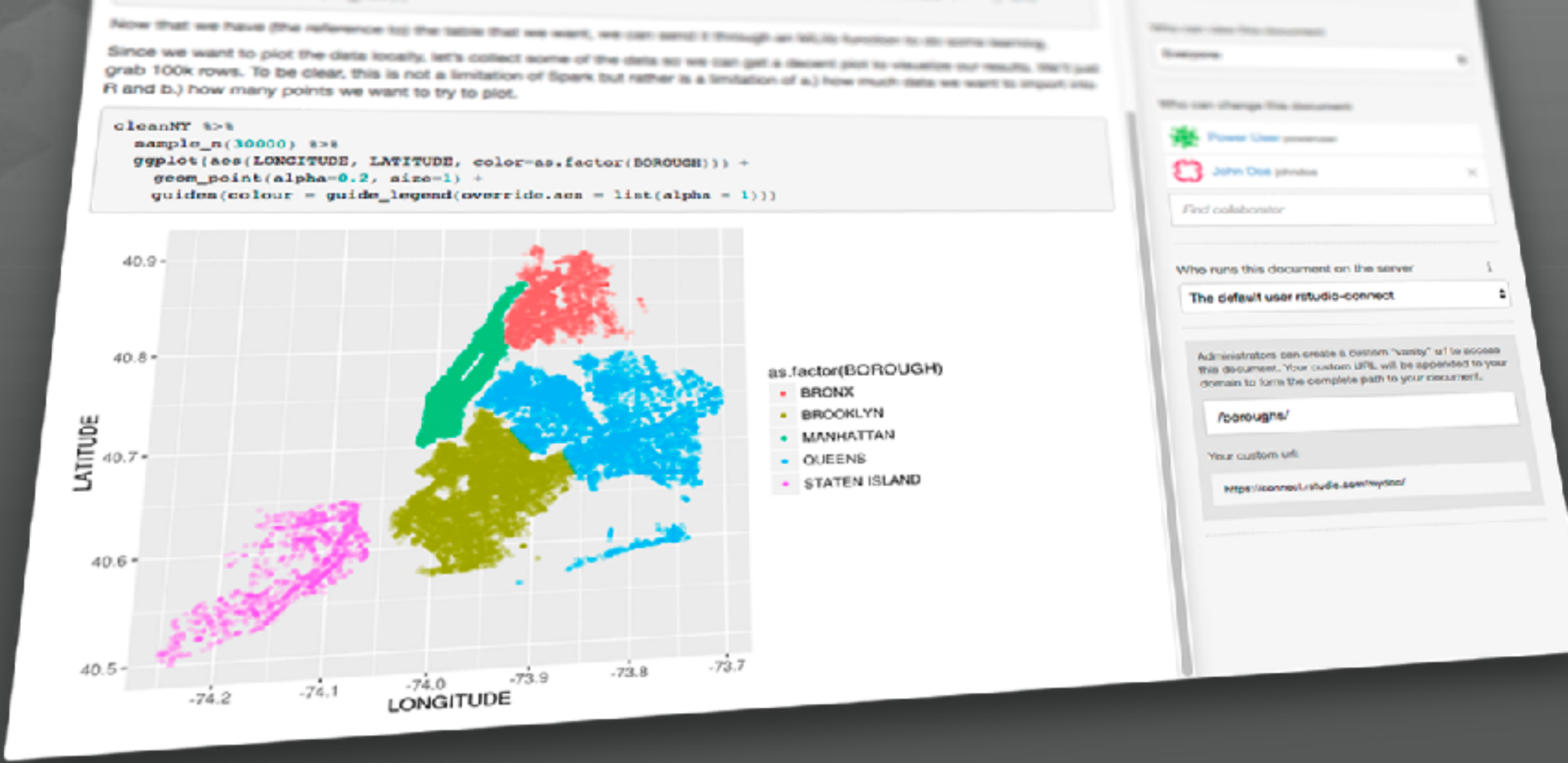
A bit of data cleaning to ensure that we only get rows with that includes the fields that we're interested in.

```
cleanNY <- nypd  %>%
  filter(!is.na(LATITUDE), !is.na(LONGITUDE), LATITUDE != 0, LONGITUDE != 0, BOROUGH != "")
 %>%
  arrange(desc(UNIQUE_KEY))

if (params$borough != "All"){
  cleanNY <- filter(cleanNY, BOROUGH == toupper(params$borough))
}
```

Now that we have (the reference to) the table that we want, we can send it through an MLlib function to do some learning.

```
cleanNY %>%
```



Now that we have (the reference to) the table that we want, we can send it through an MLlib function to do some learning.

Since we want to plot the data locally, let's collect some of the data so we can get a decent plot to visualize our results. We'll just grab 100k rows. To be clear, this is not a limitation of Spark but rather is a limitation of a.) how much data we want to import into R and b.) how many points we want to try to plot.

```
cleanNY %>%
  sample_n(30000) %>%
  ggplot(aes(LONGITUDE, LATITUDE, color=as.factor(BOROUGH))) +
    geom_point(alpha=0.2, size=1) +
   guides(colour = guide_legend(override.aes = list(alpha = 1)))
```

Who runs this document on the server

The default user rstudio-connect

Administrators can create a custom "vanity" url to access this document. Your custom URL will be appended to your domain to form the complete path to your document.

/boroughs/

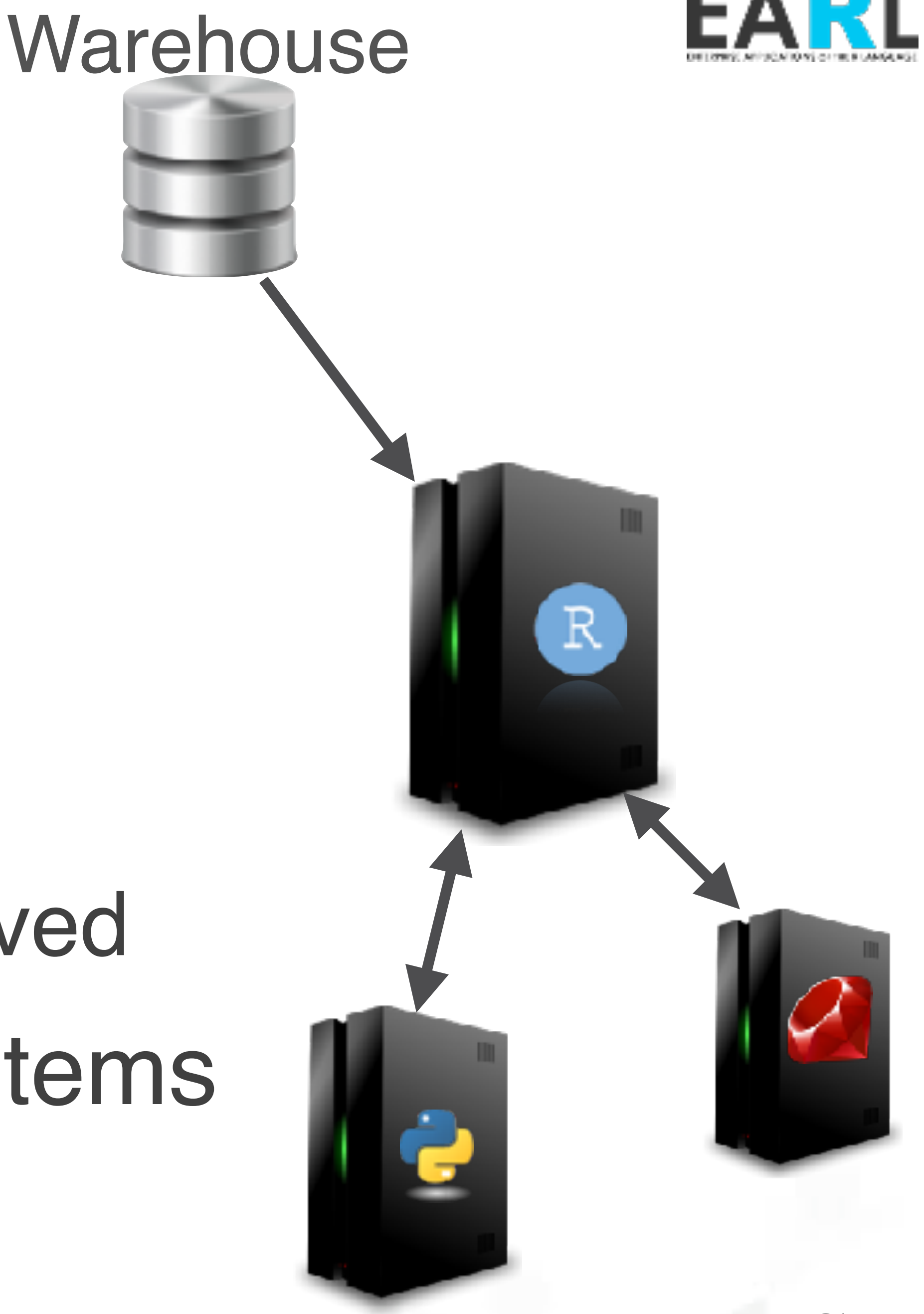
Your custom url:

https://connect.rstudio.com/myloc/

Stage 3: Fully Integrated

Stage 3: Overview

- R is fully embraced
  - Permissive access to data
  - Share analysis broadly
  - Non-data-scientists get involved
- Incorporate R into other systems



Stage 3: Issues

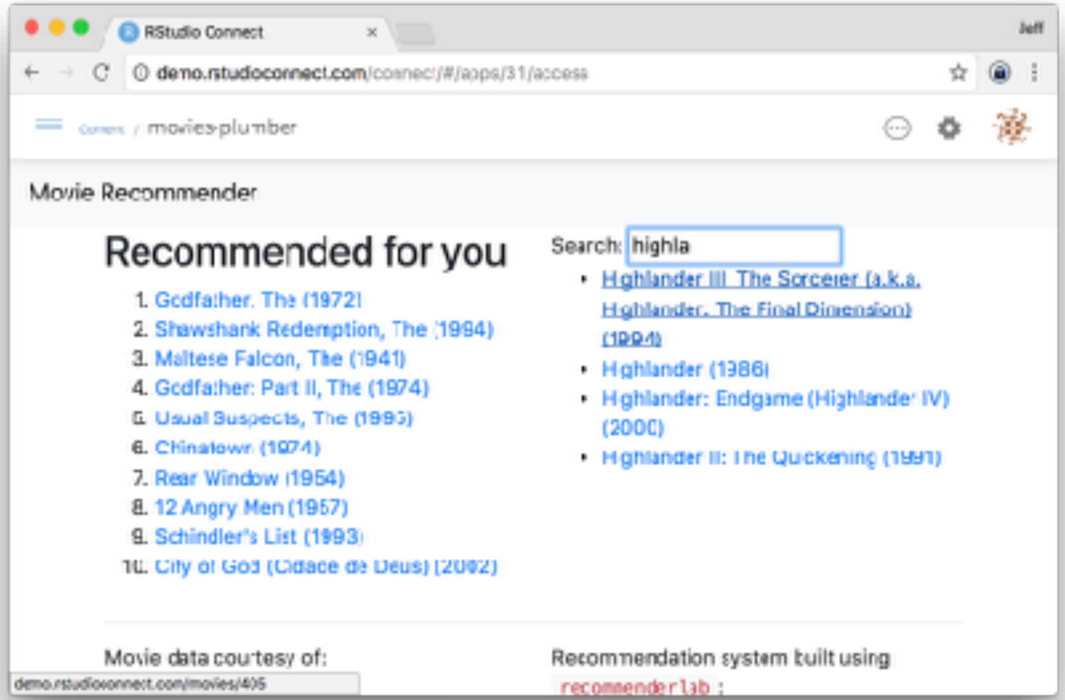
- Enabling other systems to communicate with R
- Code sharing, package management
- Data governance

Plumber

- Open-source R package
- Make your R code available over a web API
- Consumed by systems in any other programming language
- https://www.rplumber.io/

Stage 3: Use Cases

- Leverage R from existing enterprise system: Java, Python, Ruby, etc.
  - Compare to porting your R code into another language
- Recommendation/forecasting engine
- Python ETL jobs in a scheduled RMD


RStudio Connect

demo.rstudioconnect.com/connect/#/apps/31/access

Content / movies-plumber

Movie Recommender

Recommended for you

1. Godfather, The (1972)
2. Shawshank Redemption, The (1994)
3. Maltese Falcon, The (1941)
4. Godfather: Part II, The (1974)
5. Usual Suspects, The (1995)
6. Chinatown (1974)
7. Rear Window (1954)
8. 12 Angry Men (1957)
9. Schindler's List (1993)
10. City of God (Cidade de Deus) (2002)

Search: highla

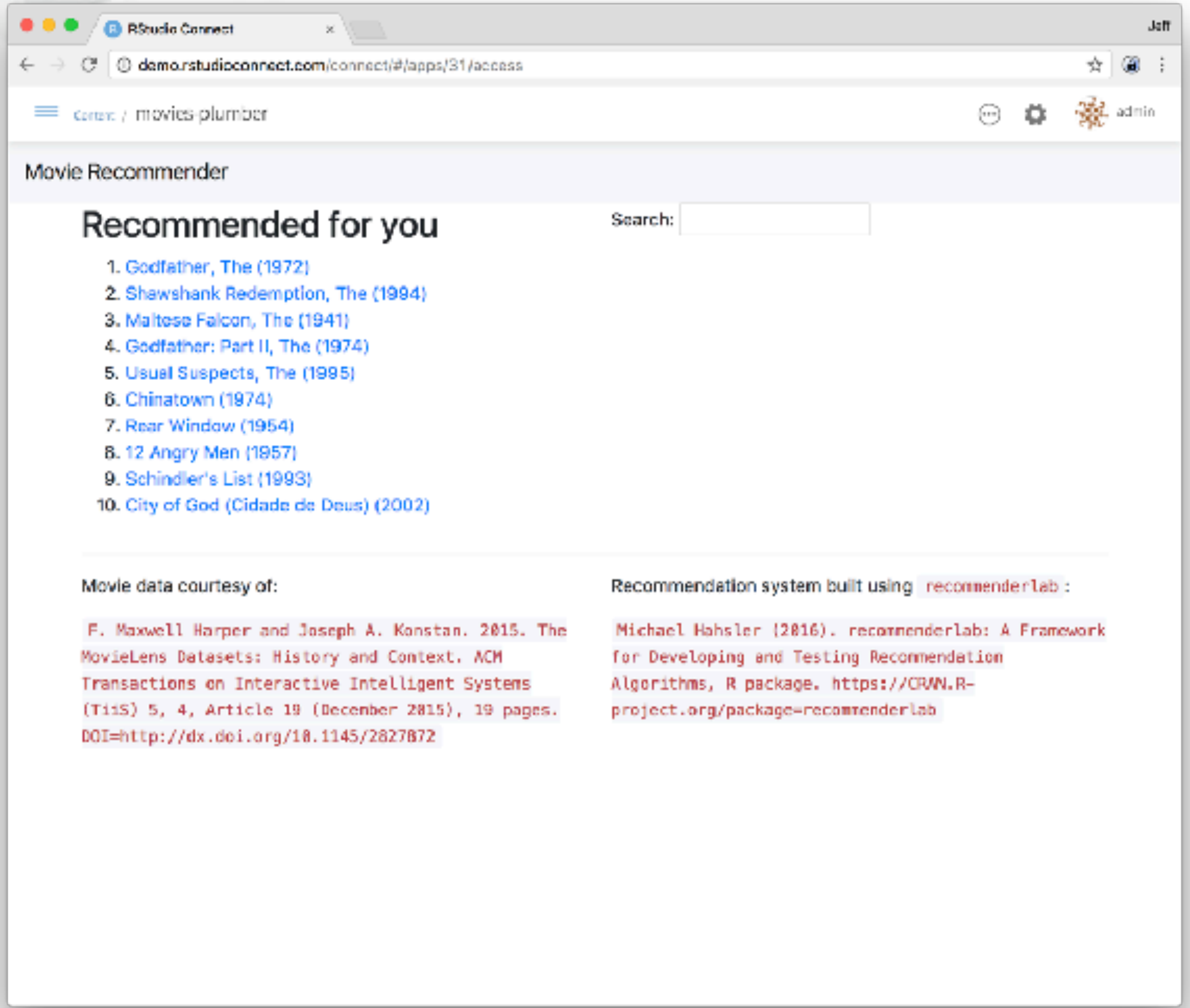
- Highlander III: The Sorcerer (a.k.a. Highlander: The Final Dimension) (1994)
- Highlander (1986)
- Highlander: Endgame (Highlander IV) (2000)
- Highlander II: The Quickening (1991)

Movie data courtesy of:

Recommendation system built using `recommenderlab`;

demo.rstudioconnect.com/movies/405

Stage 3 Demo



Movie Recommender

Recommended for you

Search:

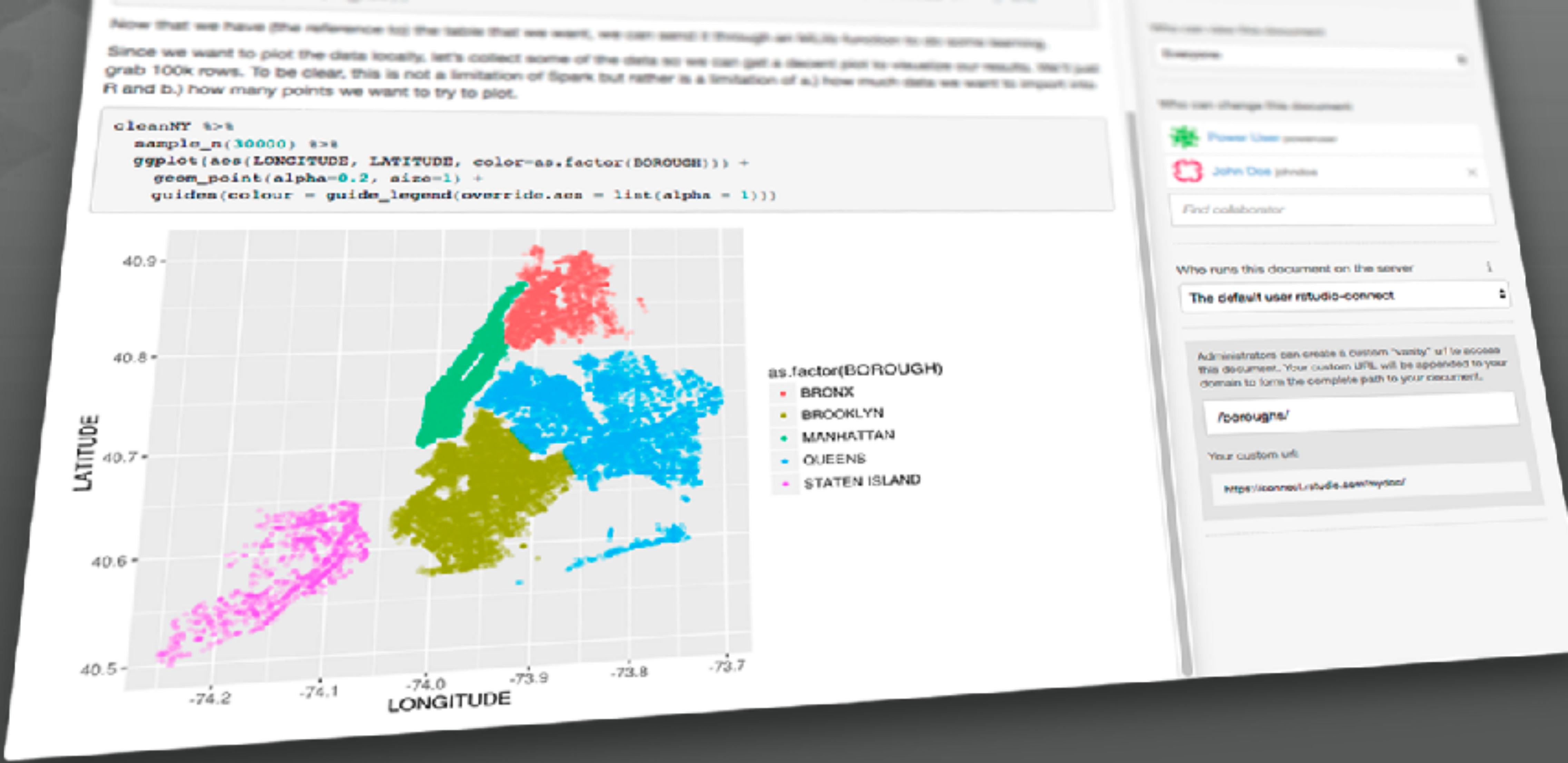
1. Godfather, The (1972)
2. Shawshank Redemption, The (1994)
3. Maltese Falcon, The (1941)
4. Godfather: Part II, The (1974)
5. Usual Suspects, The (1995)
6. Chinatown (1974)
7. Rear Window (1954)
8. 12 Angry Men (1957)
9. Schindler's List (1993)
10. City of God (Cidade de Deus) (2002)

Movie data courtesy of:

`F. Maxwell Harper and Joseph A. Konstan. 2015. The MovieLens Datasets: History and Context. ACM Transactions on Interactive Intelligent Systems (TiiS) 5, 4, Article 19 (December 2015), 19 pages. DOI=http://dx.doi.org/10.1145/2827872`

Recommendation system built using `recommenderlab`:

`Michael Hahsler (2016). recommenderlab: A Framework for Developing and Testing Recommendation Algorithms, R package. https://CRAN.R-project.org/package=recommenderlab`


Now that we have (the reference to) the table that we want, we can send it through an MLlib function to do some learning.

Since we want to plot the data locally, let's collect some of the data so we can get a decent plot to visualize our results. We'll just grab 100k rows. To be clear, this is not a limitation of Spark but rather is a limitation of a.) how much data we want to import into R and b.) how many points we want to try to plot.

```
cleanNY %>%
  sample_n(30000) %>%
  ggplot(aes(LONGITUDE, LATITUDE, color=as.factor(BOROUGH))) +
    geom_point(alpha=0.2, size=1) +
   guides(colour = guide_legend(override.aes = list(alpha = 1)))
```

Who can view this document

Everyone

Who can change this document

Power User poweruser

John Doe johndoe

Find collaborator

Who runs this document on the server

The default user rstudio-connect

Administrators can create a custom "vanity" url to access this document. Your custom URL will be appended to your domain to form the complete path to your document.

/boroughs/

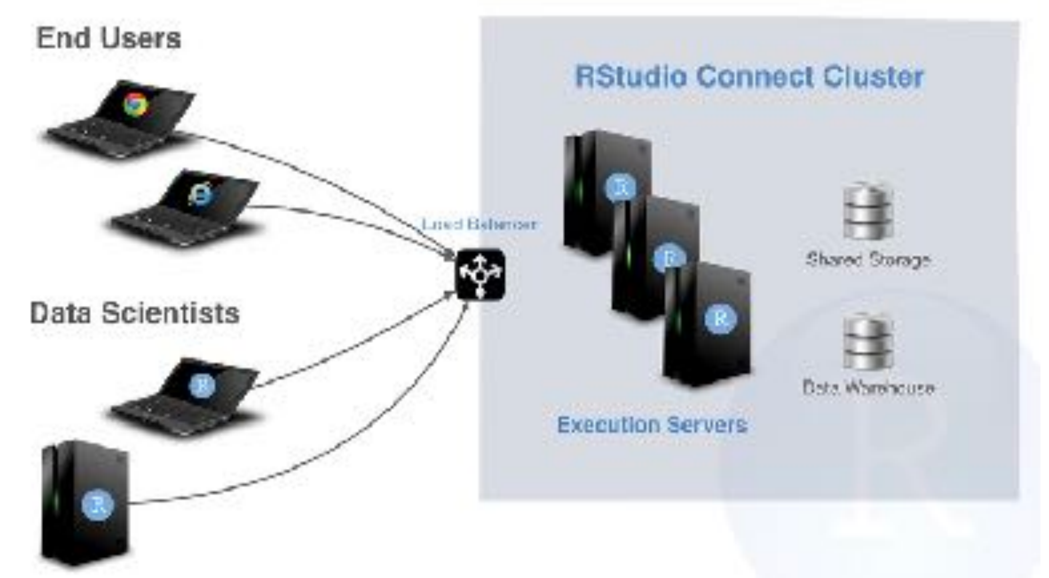
Your custom url:

https://connect.rstudio.com/mydoc/

Updates

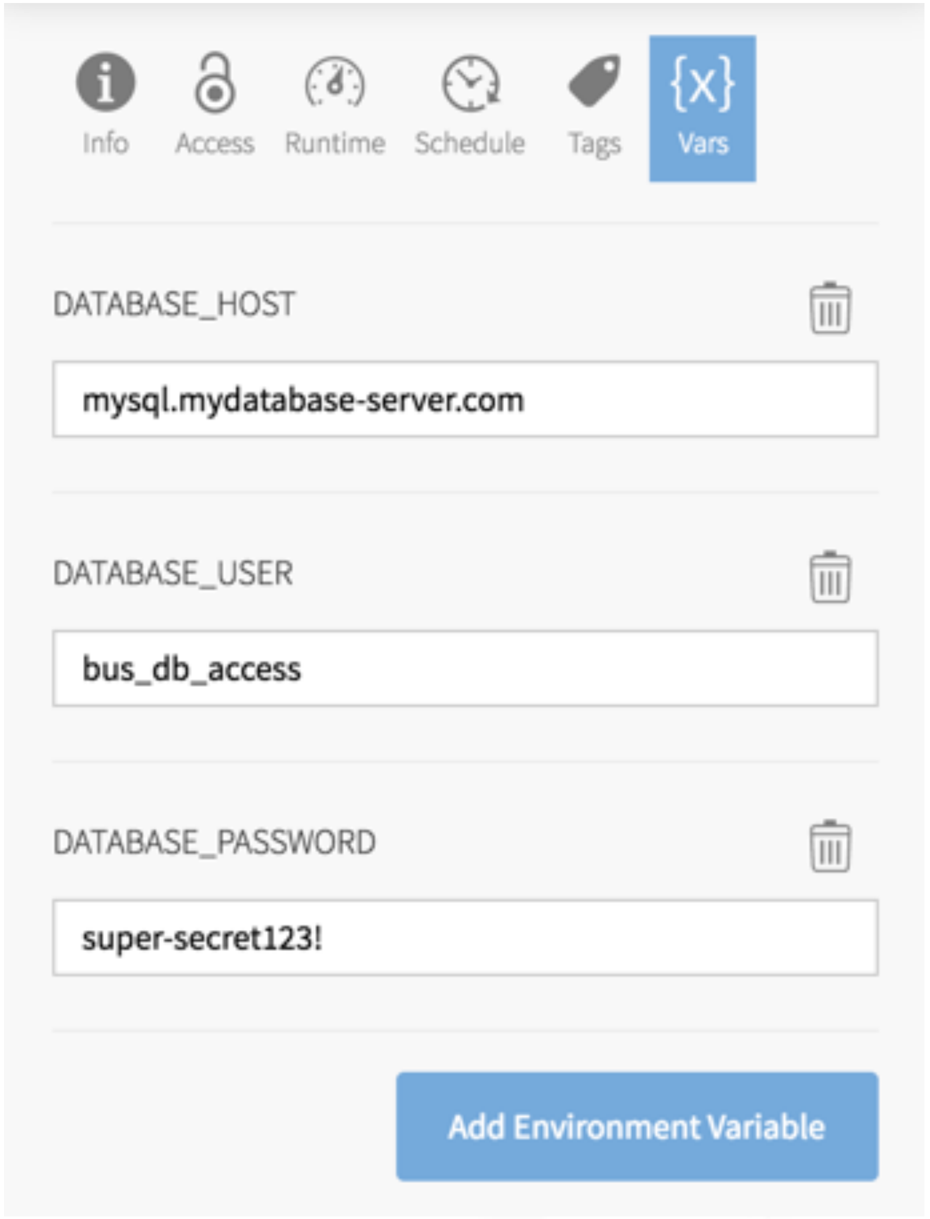
What's New?

- Load balancing & high availability
- Floating licenses
- Kerberos
- Content collaboration & continuous integration



Upcoming…

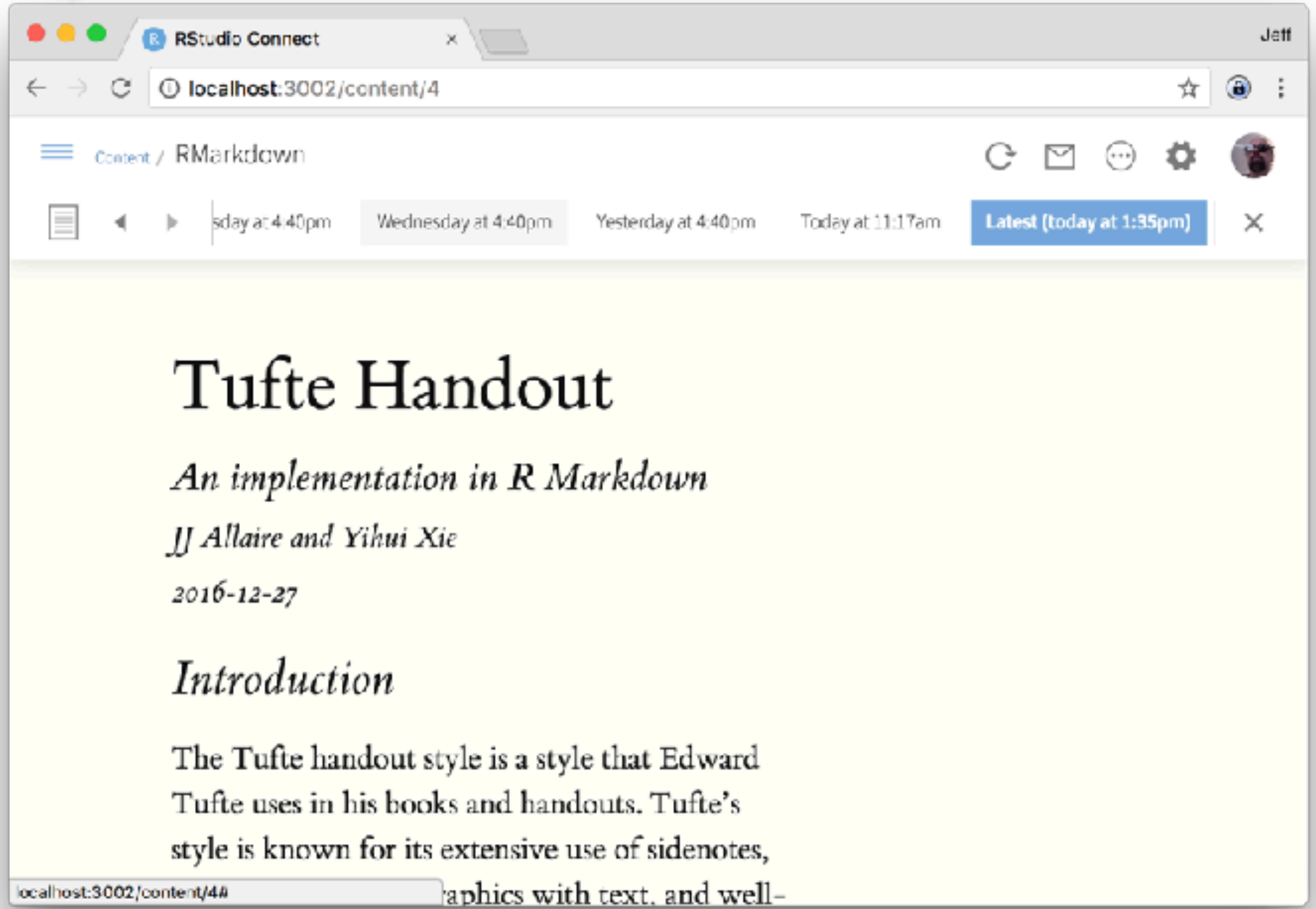
- View old versions of a report
- API to control RStudio Connect
- In-UI Management of Environment Variables
- SUSE support
- App-specific event logs



DATABASE_HOST
mysql.mydatabase-server.com

DATABASE_USER
bus_db_access

DATABASE_PASSWORD
super-secret123!



Tufte Handout

An implementation in R Markdown

JJ Allaire and Yihui Xie

2016-12-27

Introduction

The Tufte handout style is a style that Edward Tufte uses in his books and handouts. Tufte's style is known for its extensive use of sidenotes, aphics with text, and well-

Questions?

- Download & 45-day free trial: http://rstd.io/rsc
- Admin Guide: http://rstd.io/rsc-admin
- Slides: http://rstd.io/earl2017