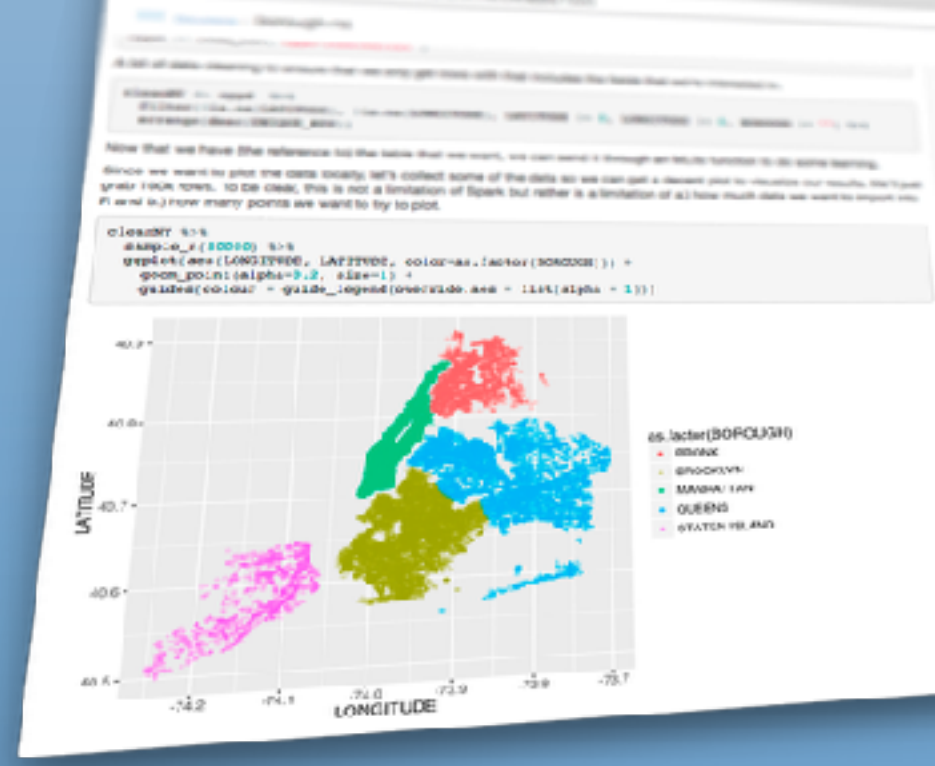
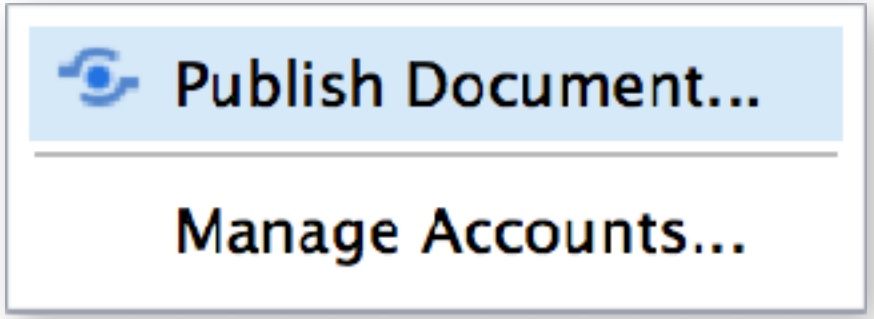
RStudio Connect Internals — The Basics

Jeff Allen 3/2017

1. What is RStudio Connect?
2. Installation & setup
3. Details & usage



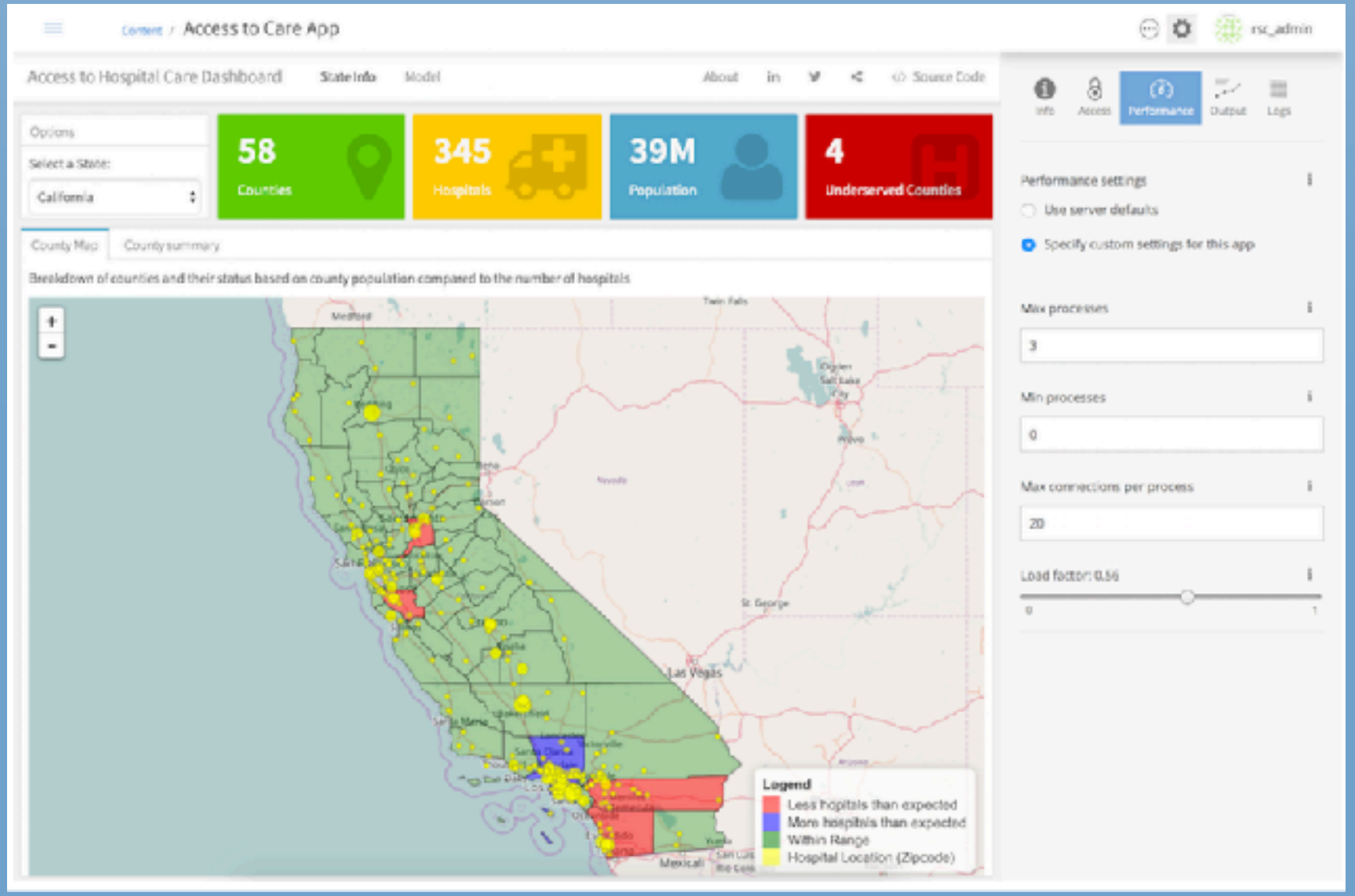
RStudio Connect



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

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

RStudio Connect Web Application



Download & Install

- Download from http://rstd.io/rsc
- Install R
  - Package manager or compile from source
- Install Connect from `.deb` or `.rpm` file
  - Ubuntu 12.04, 14.04 & 16.04
  - RedHat/CentOS 6 & 7

Configuration & Logs

```
3   [Server]
4   SenderEmail = noreply@myorg.com
5   Address = http://my.rstudioconnect.com
6
7   [Http]
8   Listen = :3939
9
10  [Https]
11  Listen = :443
12  Key = /connect/ssl/connect.key
13  Certificate = /connect/ssl/connect.crt
14
15  [Authentication]
16  Provider = password
17
18  [Database]
```

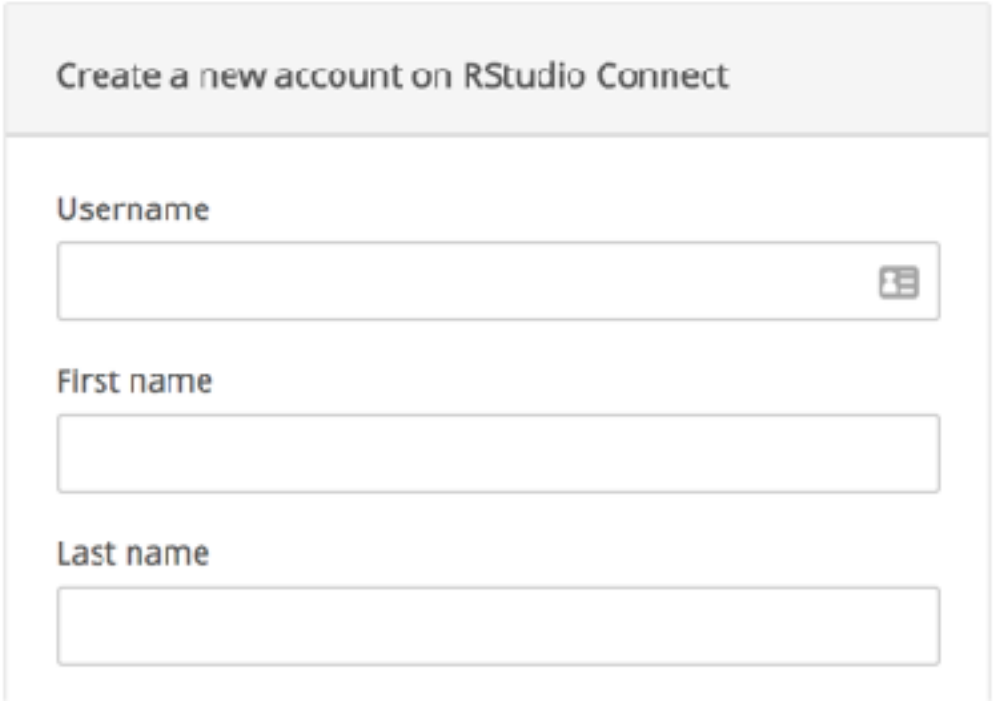
- http://rstd.io/rsc-cfg
- Many things managed in the config file at /etc/rstudio-connect/rstudio-connect.gcfg
- /var/log/rstudio-connect.log

Authentication

Authentication Method	Supported	Grant access to content to users who have not previously logged in	Group access to content
Connect's built-in authentication	●	●	●
LDAP and Active Directory – Single Bind	●		
LDAP and Active Directory – Double Bind	●	●	●
Google OAuth 2.0	●	●	
PAM	●		
Proxied Authentication	●		

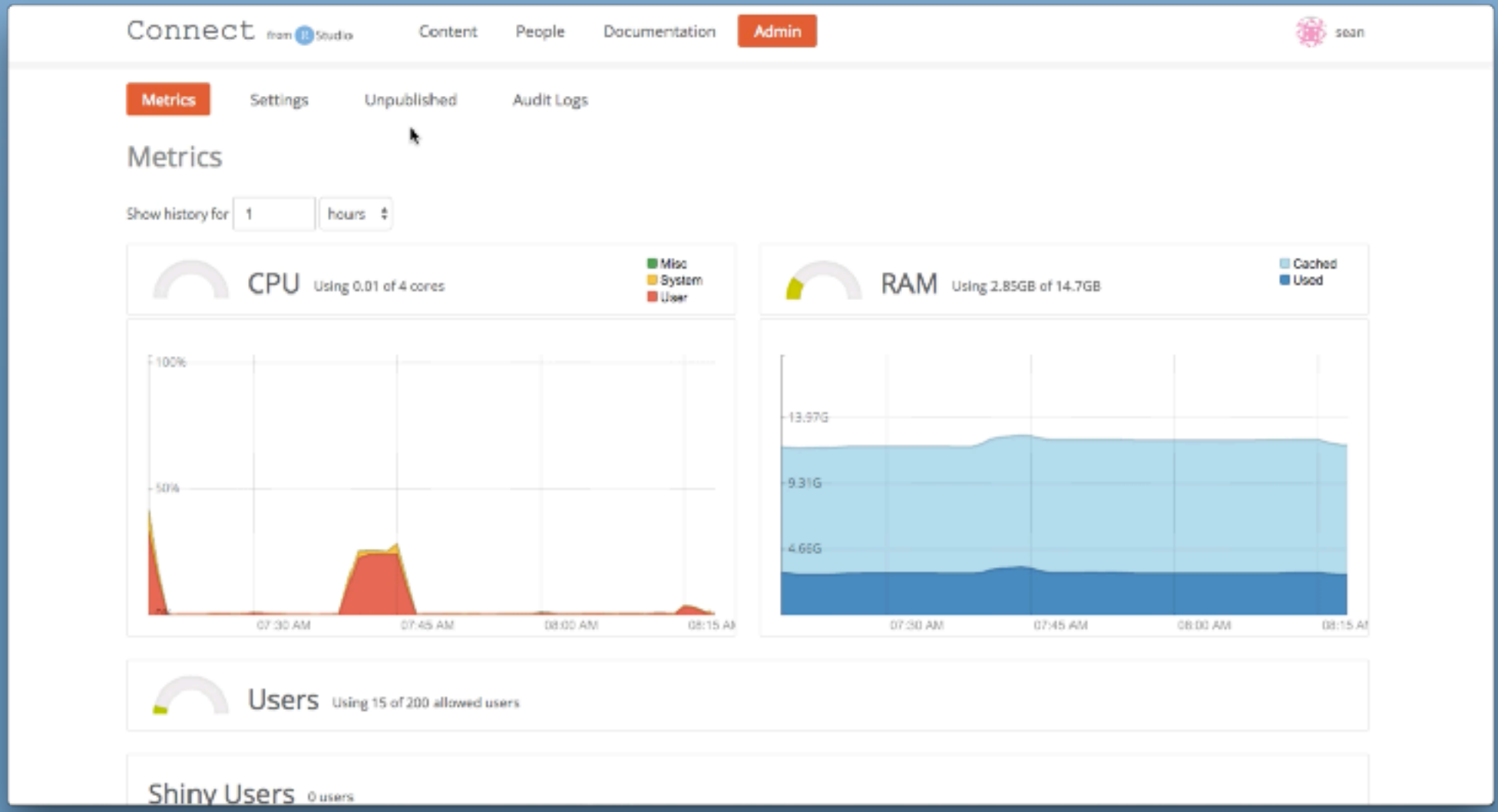
Create Your First Account

- Create an account or login using existing credentials
- The first account created becomes an admin



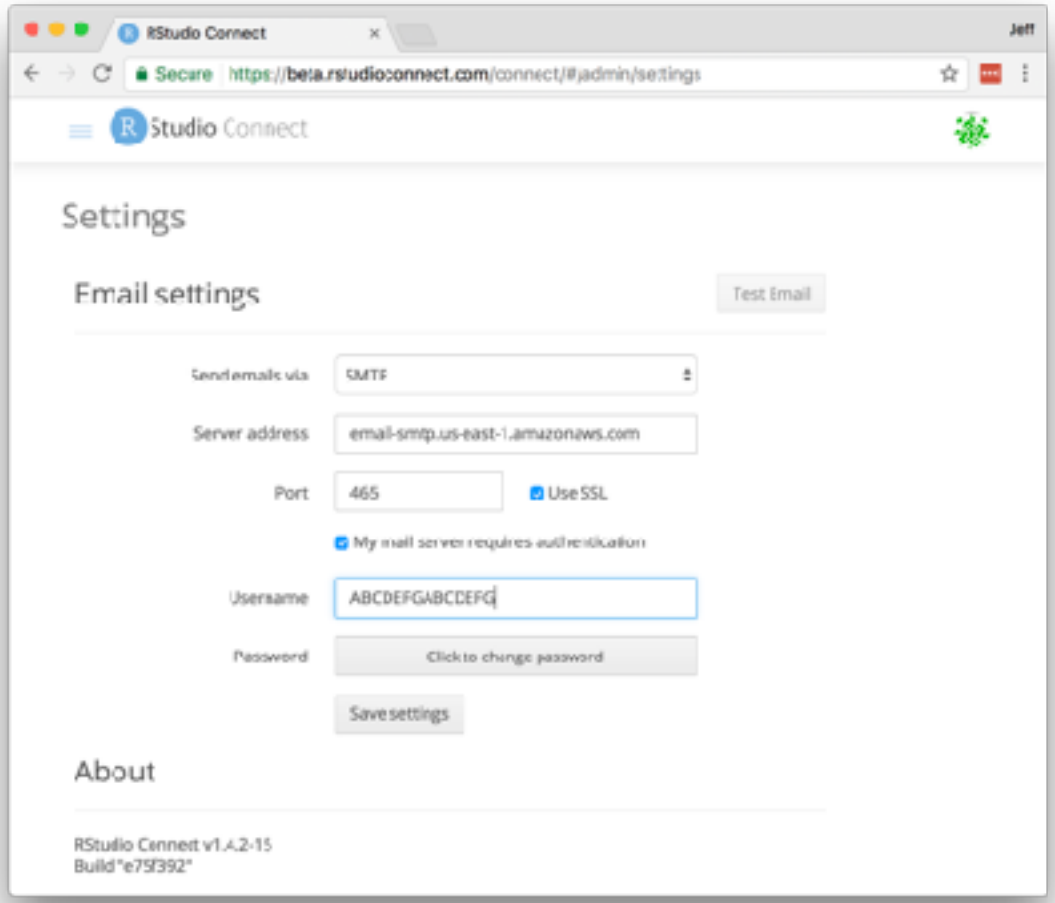
RStudio

RStudio Connect Dashboard - Admin



Configure Email

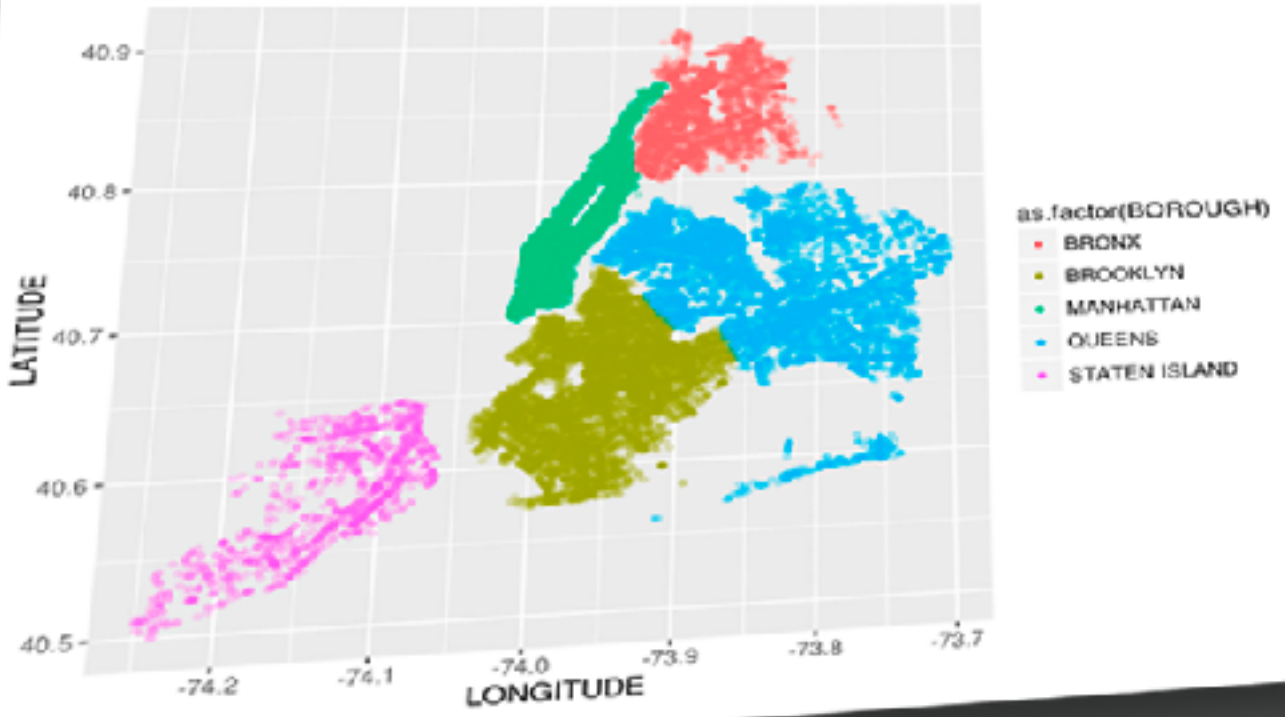
- Used for
  - Account confirmation*
  - Delivery of reports
  - Notification of errors
- Admin > Settings
  - Sendmail or an SMTP server



*Only for built-in auth

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(100000) %>%
  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:

It's alive!

Under the Hood

- Server w/ HTTP API
- Web application front-end
- Enterprise authentication
- R execution engine
- Database*

http://rstd.io/rsc-admin

RStudio

*As of early 2017, SQLite only; others coming soon.

R Execution

Spin up and shut down R processes

- On-demand
- On a schedule

Sandbox environment

- Uses “unshare” with per-process “bind mounts” to isolate user data/code from other applications
- Requires root

R Execution

R package dependencies captured and provisioned per app

- System dependencies may be required

R is inherited from your server

- We do support multiple versions of R and will try to align what we find on the client
- http://rstd.io/rsc-installing-r

Data Storage

- `/var/lib/rstudio-connect/` by default
- Recommend 50GB+
- SQLite database must be on local disk
  - All other data can be on network drive

http://rstd.io/rsc-storage

User Roles

- **Admin** - All privileges, but “irregular” actions are audited
- **Publisher** - Can upload new content.
- **Viewer** - Can see content.

Application Privileges

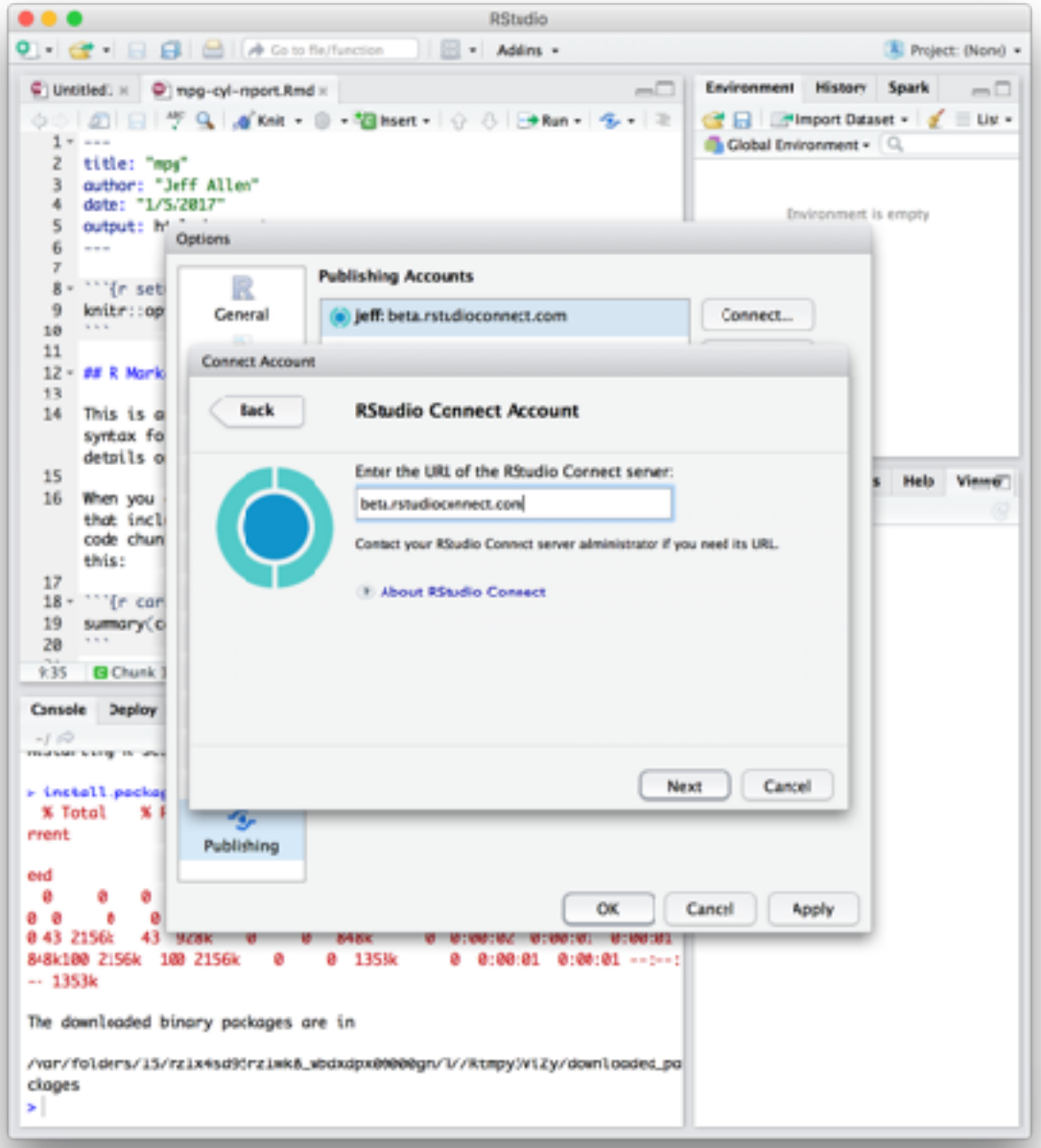
- **Collaborator** - Can publish new versions*, manage settings, refresh.
- **Viewer** - Can only see the existing content.

RStudio

*Enhanced experience coming soon.

IDE Integration

- Built on top of the rsconnect R package
- You can do all the same things via CLI using rsconnect
- In IDE: Tools > Global Options > Publishing
- Hits the same HTTP API
- Beware of custom CAs



Shiny Applications

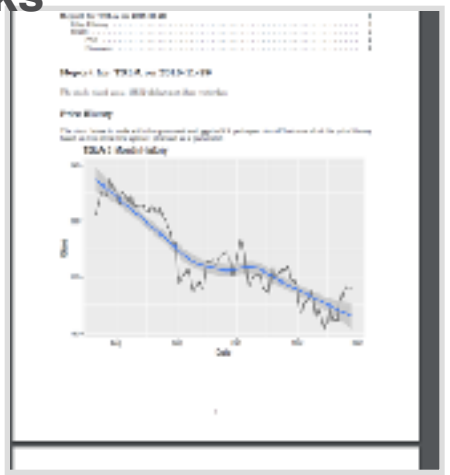
- Shiny is an open-source R package
- Requires an interactive R runtime, proxies traffic back and forth
- Connect is roughly a superset of Shiny Server
  - Authentication
  - User self-management
  - Scaling
  - Other content types

R Markdown

- Reports that interleave code and prose
- Notebooks, dashboards, storyboards, websites, books, etc.
- Render to **HTML**, PDF, Docx

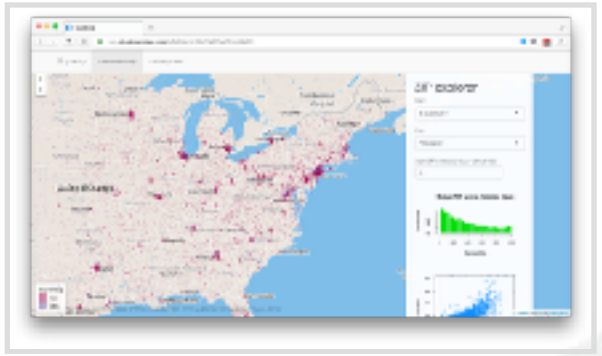
Notebooks

Dashboards

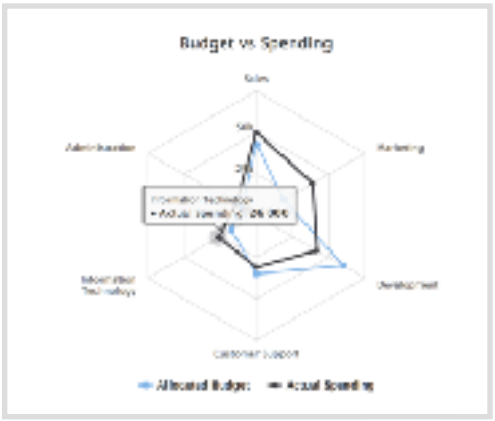


Reports

Slides



Shiny Apps

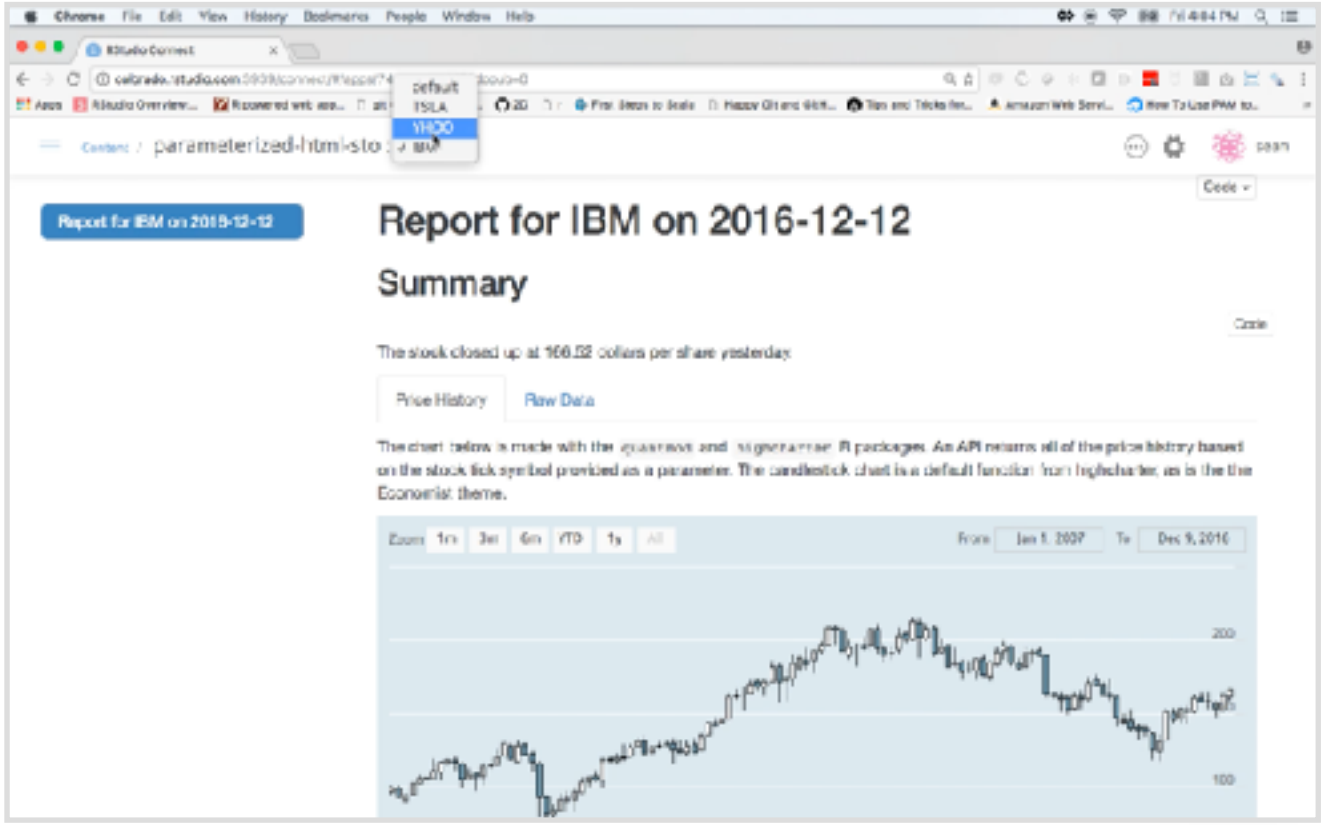


Plots

RStudio

Parameterized RMD

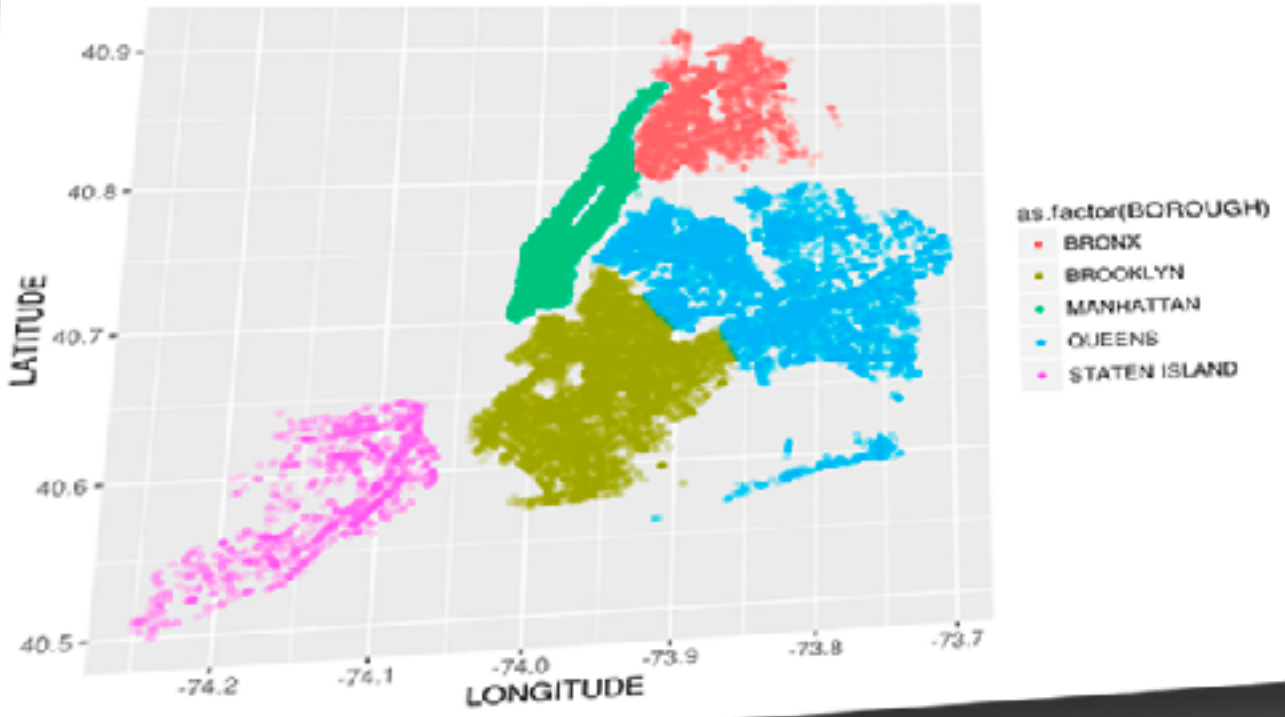
- Run a regular RMD with variable inputs
- Still output an HTML/ PDF/Docx file
- Still schedulable
- Allow others to change the inputs of the doc



RStudio

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(100000) %>%
  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:

Demo

Additional resources

- Download & 45-day free trial: http://rstd.io/rsc
- Admin Guide: http://rstd.io/rsc-admin
- IT Q&A: http://rstd.io/rsc-it-qa
- Authentication details: http://rstd.io/rsc-auth
- Release notes: http://rstd.io/rsc-news

RStudio