

# AWS and R

## Twin Cities R User Group (TCRUG) Sept. 20, 2018

LeSean Bruneau

Twitter: @leseanbruneau

LinkedIn: <http://www.linkedin.com/in/lesean-bruneau-7092625>

Github: <https://github.com/mndatascienceexamples/AWS-and-R>

Blog: <http://datascienceexamples.com>

# Agenda

- AWS Services for demo projects
- Overview demo projects
- Project 1 – RStudio Local and AWS S3
  - RStudio local integration with AWS S3
  - Run Web App with RStudio local data results
- Project 2 – RStudio Server on AWS EC2
  - Install R on AWS EC2 instance
  - Install RStudio Server on AWS EC2 instance
  - Run RStudio Server on EC2 Instance

# AWS Services

- EC2 Instance (Compute)
  - Linux server
  - Security group permissions
- IAM (Authentication and Authorization)
- S3 Bucket (Storage)
  - Directories and Files
  - Permissions
    - Assign permissions on files/directories
    - Use IAM for program access to S3 Bucket

# Project 1 Demo Overview

- All MLB 2017 Regular Season Games
  - <http://baseball-reference.com>
- S3 Bucket for data input file
- RStudio Local
  - Create dataframe from data input file in S3
  - R Function to select one team's games
  - Output – write JSON file to S3 Bucket
- S3 Bucket for serverless web application
  - Display output from RStudio desktop on web page

# Project 2 Demo Overview

- Create AWS EC2 Instance
- Install R and R system packages
- Install RStudio Server
- Run RStudio Server on a web browser

# Project Setup Information - Local

- Create local workspace
  - Demo: OS: Windows; Directory: C:\R directory
- Github Repo
  - <https://github.com/mndatascienceexamples/AWS-and-R>
  - Extract zip file to local workspace (C:\R\AWS-and-R)

# Project Setup Information - AWS

- **AWS IAM User Account**

- Access Type: Programmatic access
- Policy Name: PowerUserAccess
- Save Secret Key and Access Key

- **AWS S3 Bucket**

- Create S3 Bucket
- Upload Github R directory (c:\R\AWS-and-R\R):
  - AWS S3 upload default options
- Upload Github webapp directory (c:\R\AWS-and-R\webapp):
  - Public Read-only directory access
  - All other AWS S3 upload default options

# Project 1 – RStudio and S3 Setup

- RStudio Desktop
  - Set working directory – C:\R
  - Set System Env Variables for AWS IAM Account
    - `Sys.setenv("AWS_ACCESS_KEY_ID" = "<PUT-ACCESS-KEY>", "AWS_SECRET_ACCESS_KEY" = "<PUT-SECRET-KEY>")`
- Install R Libraries
  - `> Install_Libraries.R`
- Load libraries and create R function
  - `> Create separate function`
- Upload webapp directory to S3 Bucket
  - Note: Grant Public Read-only permission to webapp directory and files



# Project 1 – RStudio and S3

- Load data and create header names
- Execute R Function with Team Name Abbr.
- Write results to JSON file
- Upload JSON file to S3
- Verify results in web application

# Project 2 – RStudio Server EC2 Setup

- Create EC2 Instance

- AMI: Amazon Linux 2 AMI (HVM), SSD Volume Type
- Instance Type: General Purpose t2.micro
- Configuration Instance: <<default options>>
- Add storage: <<default options>>
- Add tags: {Key: Name; Value: R Compute Server}
- Configure Security Group – Create group opening following ports
  - SSH – Port Range: Port 22 – Source: <YOUR\_IP\_ADDRESS>/32
  - Custom TCP Rule – Port Range: Port 8787 – Source: 0.0.0.0/0
- Review: <<default options>>; Launch Instance
- Key Pair: New or existing

# Project 2 – RStudio Server EC2 Setup

- EC2 Instance Update
  - > sudo yum -y update
- EC2 Instance – Install Git
  - > sudo yum -y install git
- EC2 Instance – Clone Git Repo
  - From ec2-user home directory (/home/ec2-user)
  - > git clone <https://github.com/mndatascienceexamples/AWS-and-R>

# Project 2 – RStudio Server Install

## Four Scripts for R, RStudio Server installation

(/home/ec2-user/AWS-and-R/R/server)

- **Script1\_install\_ec2\_utils\_sudo.sh**
  - Install dependencies for R installation on Linux server
- **Script2\_download\_R\_utils.sh**
  - Download R installation package and configure installation
- **Script3\_install\_R\_sudo.sh**
  - R installation on Linux server
- **Script4\_install\_R\_packages\_sudo.sh**
  - R system libraries installation on Linux server

# Project 2 – RStudio Server User

Create user on Linux server for RStudio Server

- > adduser rstudio
- > sh -c “echo rstudio | passwd rstudio -stdin”

# Project 2 – RStudio Server Connect

## Web Browser – Connect to RStudio Server

- `http://<EC2_INSTANCE_IP_ADDRESS>:8787`
- Login with Linux user `rstudio` and password
- Check RStudio Server working directory