



R at Microsoft

David Smith
@revodavid



About me

Cloud Advocate, Microsoft
Director, R Consortium
Co-author, *Introduction to R*
Editor, blog.revolutionanalytics.com

davidsmi@microsoft.com
[@revodavid](https://twitter.com/revodavid)



"Bit" art by @ashleymcnamara

2007: Revolution Computing founded



2009-01-06: R hits the Front Page*

The New York Times

Business Computing

WORLD U.S. N.Y. / REGION BUSINESS TECHNOLOGY SCIENCE HEALTH SPORTS OPINION

Data Analysts Captivated by R's Power



Left, Stuart Isett for The New York Times; right, Kieran Scott for The New York Times

R first appeared in 1996, when the statistics professors Robert Gentleman, left, and Ross Ihaka released the code as a free software package.

By ASHLEE VANCE
Published: January 6, 2009

*of the technology section of the NYT

2015-04-06: Microsoft acquires Revolution Analytics

 Microsoft | Official Microsoft Blog

Microsoft to acquire Revolution Analytics to help customers find big data value with advanced statistical analysis

Jan 23, 2015 | [Joseph Sirosh - Corporate Vice President, Data Group, Microsoft](#)

[!\[\]\(7e0d6a31a51eb3952a6a6daebf7e401c_img.jpg\)](#) [!\[\]\(67f2870f14660d0c1a2612f703dff40d_img.jpg\)](#) [!\[\]\(f8bb5cf81f1458854fa051ea367dd4d6_img.jpg\)](#)

Update: April 6, 2015: Microsoft has closed the acquisition of Revolution Analytics. For more details, please read the blog post by Joseph Sirosh [here](#).

I'm very pleased to announce that Microsoft has reached an agreement to acquire [Revolution Analytics](#). Revolution Analytics is the leading commercial provider of software and services for R, the world's most widely used programming language for statistical computing and predictive analytics. We are making this acquisition to help more companies use the power of R and data science to unlock big data insights with advanced analytics.

As their volumes of data continually grow, organizations of all kinds around the world – financial, manufacturing, health care, retail, research – need powerful analytical models to make data-driven decisions. This requires high performance computation that is "close" to the data, and scales with the business' needs over time. At the same time, companies need to reduce the data science and analytics skills gap inside their organizations, so more employees can use and benefit from R. This acquisition is part of our effort to address these customer needs.

2015-06-30: R Consortium founded

THE LINUX FOUNDATION PROJECTS



Linux Foundation Announces R Consortium to Support Millions of Users Around the World

By | June 30, 2015 | Announcement

Data scientists from industry and academic research will work together to advance world's most popular language for analytics and data science and support the rapid growth of the R user community

SAN FRANCISCO, Calif., June 30, 2015 – The Linux Foundation, the nonprofit organization dedicated to accelerating the growth of Linux and collaborative development, today announced the R Consortium. This new organization will strengthen both the technical and user communities as a Collaborative Project hosted at Linux Foundation.

Provides funding for:

- R user groups
- R-Ladies
- R-hub
- useR! video streaming
- Community projects, events, and working groups

2020: Microsoft supports R

The screenshot shows the homepage of the Microsoft R Application Network (MRAN) at <https://mran.microsoft.com>. The page features a dark header with navigation links for Home, About R, Microsoft R Open, R Packages, R Community, and R Tools. A search bar offers a "Find an R Package" option. A prominent yellow banner at the top reads "Welcome to MRAN" on the left and "Download Microsoft R Open 4.0.2 now." on the right. Below the banner is a large image of a woman looking thoughtfully at a computer screen displaying a plot. Text on the left side of the image states: "R is the world's most powerful programming language for statistical computing, machine learning and graphics and has a thriving global community of users, developers and contributors." The main content area is divided into three sections: "Microsoft R Open" (with a monkey icon), "R Packages" (with a gear icon), and "CRAN Time Machine" (with a calendar icon). Each section includes a brief description and a "Download Now" button.

Microsoft R Application Network

[Home](#)

[About R](#)

[Microsoft R Open](#)

[R Packages](#)

[R Community](#)

[R Tools](#)

[Find an R Package](#)

Welcome to MRAN

[Download Microsoft R Open 4.0.2 now.](#)

R is the world's most powerful programming language for statistical computing, machine learning and graphics and has a thriving global community of users, developers and contributors.

Microsoft R Open

Microsoft R Open is the enhanced distribution of R from Microsoft Corporation. The current release, Microsoft R Open 4.0.2, is based the statistical language R-4.0.2 and includes additional capabilities for improved performance, reproducibility and platform support.

[Download Now](#)

R Packages

[Packages](#) extend R with new function and data. Whether you're using R to optimize portfolio, analyze genomic sequences, or to predict component failure times, experts in every domain have made resources, applications and code available for free online.

[Explore Packages](#)

CRAN Time Machine

For the purpose of [reproducibility](#), MRAN hosts [daily snapshots](#) of the CRAN R packages and R releases as far back as Sept. 17, 2014.

Use our [Time Machine](#) to browse CRAN contents from the past.

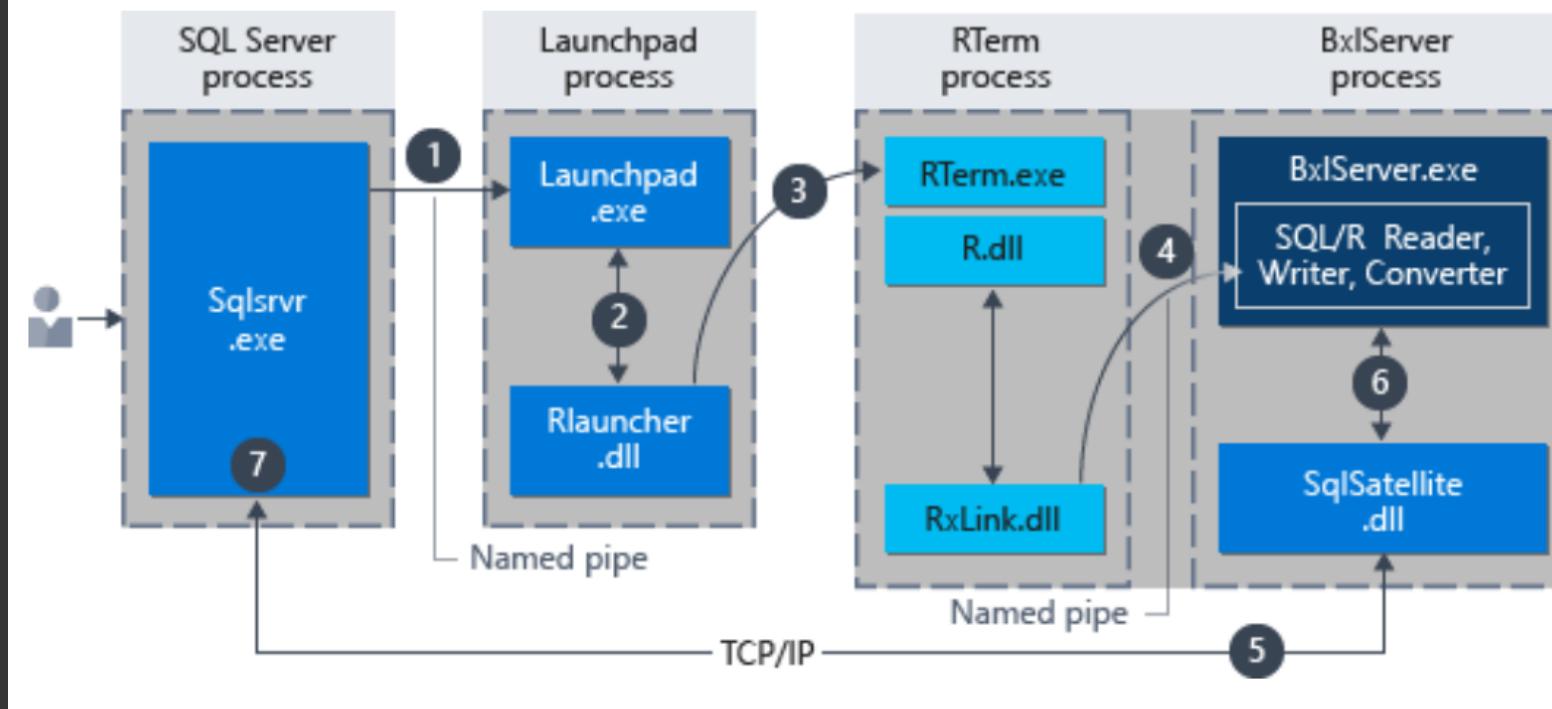
[Browse Snapshots](#)

github.com/revodavid/R-at-Microsoft

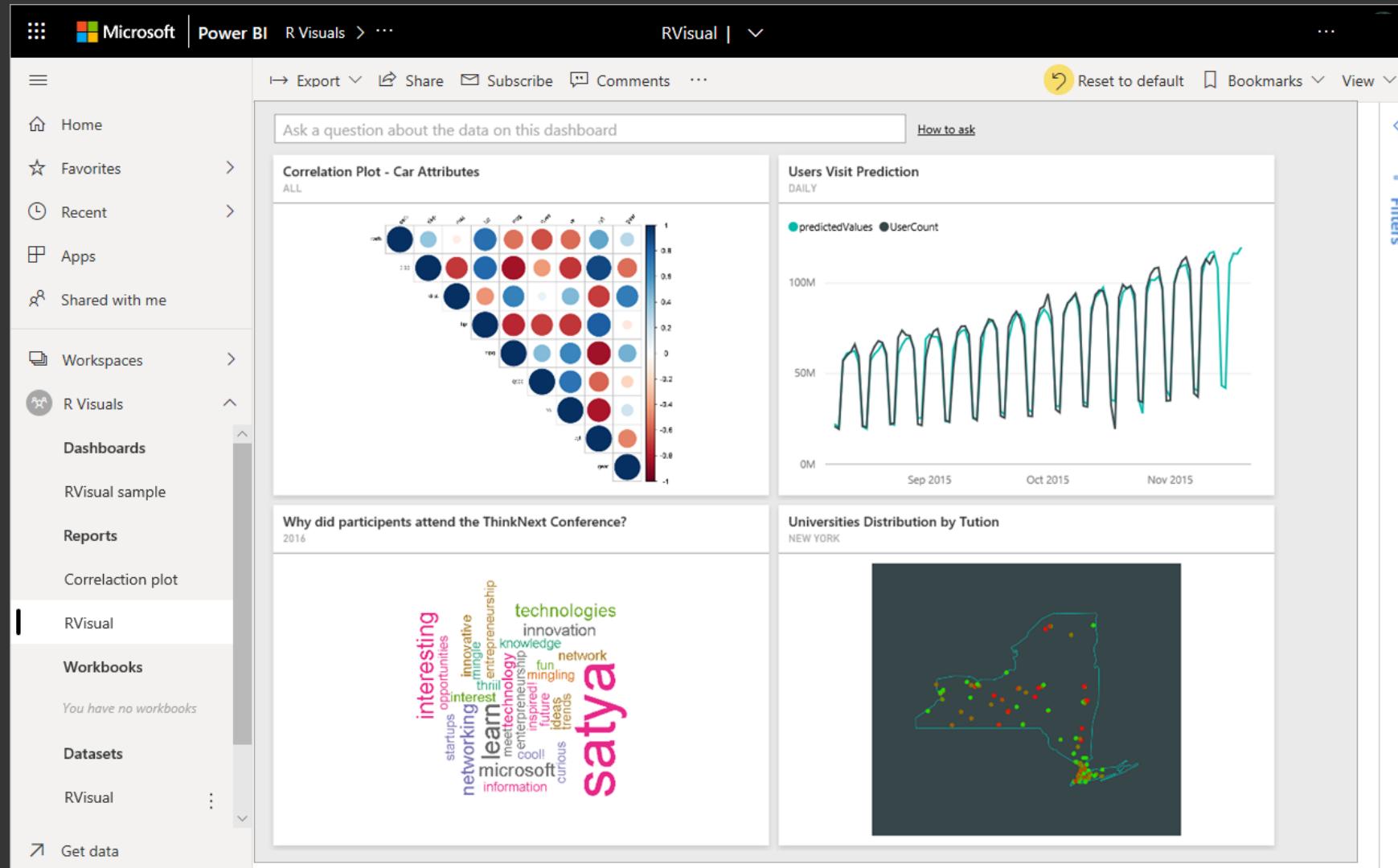
@revodavid at #whyR

R in SQL Server

R code that is run from "inside" SQL Server is executed by calling a stored procedure. Thus, any application that can make a stored procedure call can initiate execution of R code. Thereafter SQL Server manages the execution of R code as summarized in the following diagram.



R in Power BI



R in Azure

aka.ms/R-in-Azure

[Azure DataBricks](#)

[Azure HDInsight](#)

[Azure Data Science VM](#)

[Azure ML Service](#)

Microsoft | Docs Documentation Learn Q&A Code Samples

Azure Product documentation ▾ Architecture ▾ Learn Azure ▾ Develop ▾ Resources ▾

Azure / Architecture / Data Architecture Guide

R developer's guide to Azure

04/02/2020 • 8 minutes to read • 5 people +8

Many data scientists dealing with ever-increasing volumes of data are looking for ways to harness the power of cloud computing for their analyses. This article provides an overview of the various ways that data scientists can use their existing skills with the [R programming language](#) in Azure.



Microsoft has fully embraced the R programming language as a first-class tool for data scientists. By providing many different options for R developers to run their code in Azure, the company is enabling data scientists to extend their data science workloads into the cloud when tackling large-scale projects.

Let's examine the various options and the most compelling scenarios for each one.

Azure services with R language support

This article covers the following Azure services that support the R language:

Service	Description
Azure Machine Learning Server	enterprise software for data science, providing R and Python interpreters
Data Science Virtual Machine	a customized VM to use as a data science workstation or as a custom compute target
ML Services on HDInsight	cluster-based system for running R analyses on large datasets across many

AzureR: Manage Azure services from R

AzureRMR



- Lightweight, extensible R6-based interface to [Azure Resource Manager](#)
- Manage subscriptions and resource groups
- Create, update and delete [resources and templates](#)
- Work with [role-based access control \(RBAC\)](#)

AzureKeyVault



- Resource Manager and client interface to [Azure Key Vault](#)
- Secure facility for [passwords, cryptographic keys, certificates, storage account logins](#)
 - Encrypt and decrypt with keys
 - Sign and verify certificates
 - Rotate storage account access keys

AzureStor



- Resource manager and client interface to [Azure blob storage, file storage and Data Lake storage gen2](#)
- List, upload and download files
- Authenticate with access key, SAS or AAD token
- Fast [parallel transfers](#) for multiple files

AzureContainers



- Interface to [Azure Container Registry](#), [Azure Container Instances](#) and [Azure Kubernetes Service](#)
- Push and pull images to and from ACR
- Spin up containers with ACI
- Create AKS clusters and [deploy and manage services](#)

github.com/Azure/AzureR

AzureGraph



- R6-based interface to [Microsoft Graph](#)
- Emphasis on [registered apps and service principals](#), to support other packages in family
- Can be extended to work with other services in Graph: SharePoint, OneDrive, Outlook, device management, etc

AzureVM



- Flexible, configurable interface to virtual machines and [virtual machine scale sets](#)
- Customise your deployment by virtual network, security rules, IP address, load balancer, and more
- Prebuilt configurations for popular Windows and Linux images

AzureKusto



- Resource Manager and client interface to [Azure Data Explorer](#), aka Kusto
- Query data using [dplyr](#) and [DBI](#) interfaces
 - Built in the manner of dplyr (delayed execution)
- On control plane side, create and manage database principals

AzureAuth



- [OAuth authentication](#) for Azure Active Directory
- Supports [multiple authentication flows](#): authorization_code, device_code, resource_owner, client_credentials
- Supports AAD v1.0 and v2.0
- Authenticate with password or [private key \(certificate\)](#)

Azure ML Service: MLOPS and GitHub Actions

accident

Edit table Refresh Reset view Add chart

Add filter Include child runs

Run status

Status	Count
Running	0
Completed	52
Failed	3
Other	0

Show only selected rows (10 selected)

Run	Run ID	Status	Submitted time	Duration	Submitted by	Compute target
Run 55	accide...	Completed	Aug 13, 2020 3:52 PM	4m 40s	Service principal	rcluster

Accuracy

Run number	Accuracy
46	0.956
47	0.954
48	0.9545
49	0.956
50	0.955
51	0.9555
52	0.956
53	0.954
54	0.9565
55	0.957

revodavid / mlops-r-gha

Code Issues Pull requests Actions ...

All workflows New

Filter workflows

86 results

Event	Status	Branch	Actor
purple bar	Quick Deploy to Shiny #22: Commit a02ba71 pushed by revodavid	main	3 months ago 32s
blue bar	Quick Deploy to Shiny #21: Commit 850df0b pushed by revodavid	main	3 months ago 35s
purple bar	Quick Deploy to Shiny #20: Commit b083225 pushed by revodavid	main	3 months ago 30s

Accident Fatality Probability Estimator

Occupant Age:

16 69 95

Occupant gender:

f ▾

Occupant role:

pass ▾

Vehicle Year:

1,955 2,002

Seatbelt:

belted ▾

Airbag:

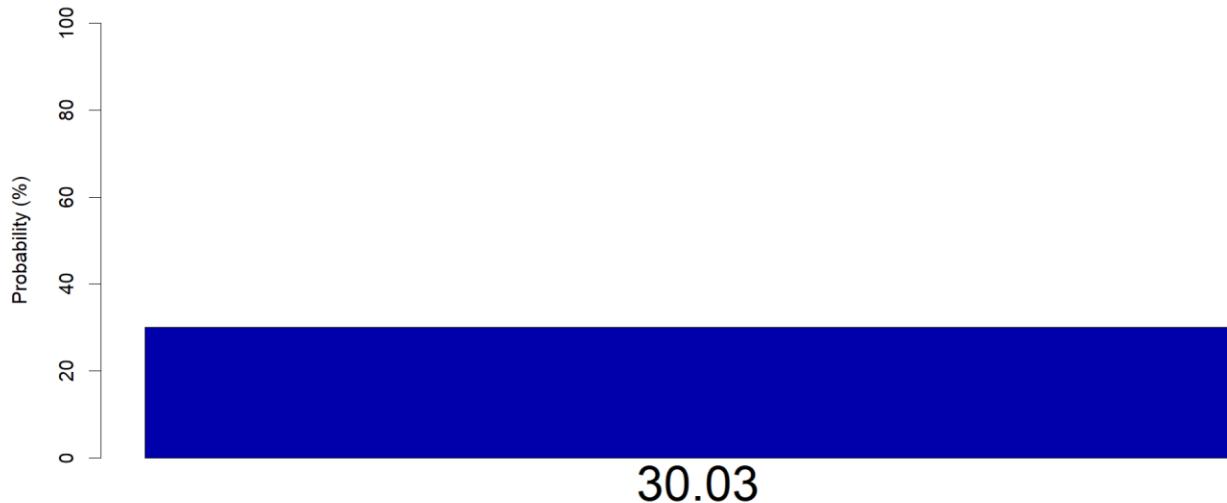
none ▾

Impact speed:

40-54 ▾

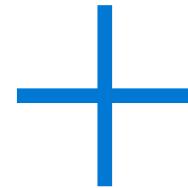
Collision type:

notfrontal ▾



Azure Machine Learning service

Set of Azure
Cloud Services



Python & R
SDKs



GitHub
Actions

That enables
you to:

- ✓ Prepare Data
- ✓ Build Models
- ✓ Train Models

- ✓ Manage Models
- ✓ Track Experiments
- ✓ Deploy Models

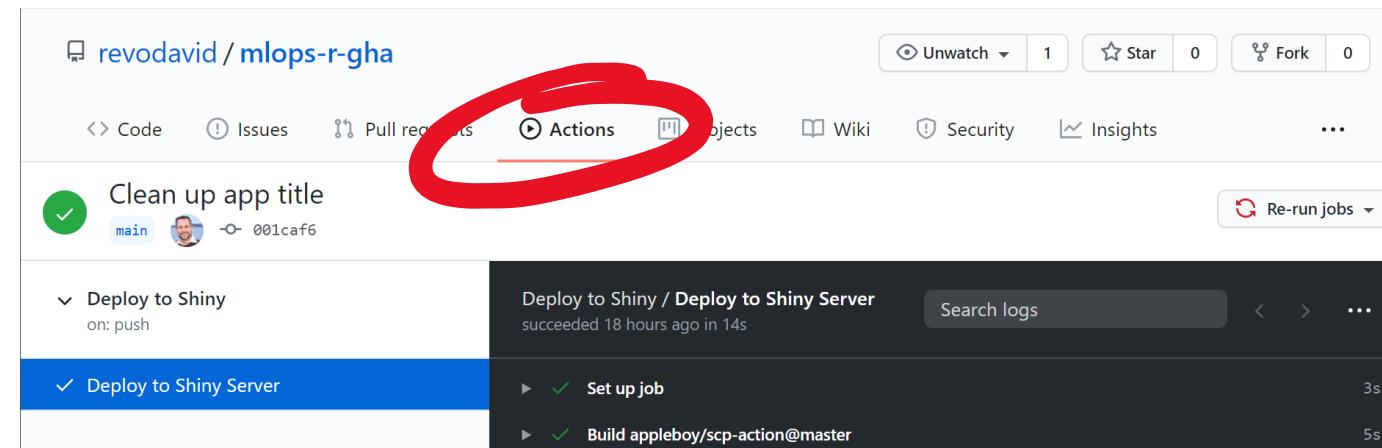
- ✓ Manage Code
- ✓ Collaborate
- ✓ Continuous Integration

GitHub Actions TL;DR

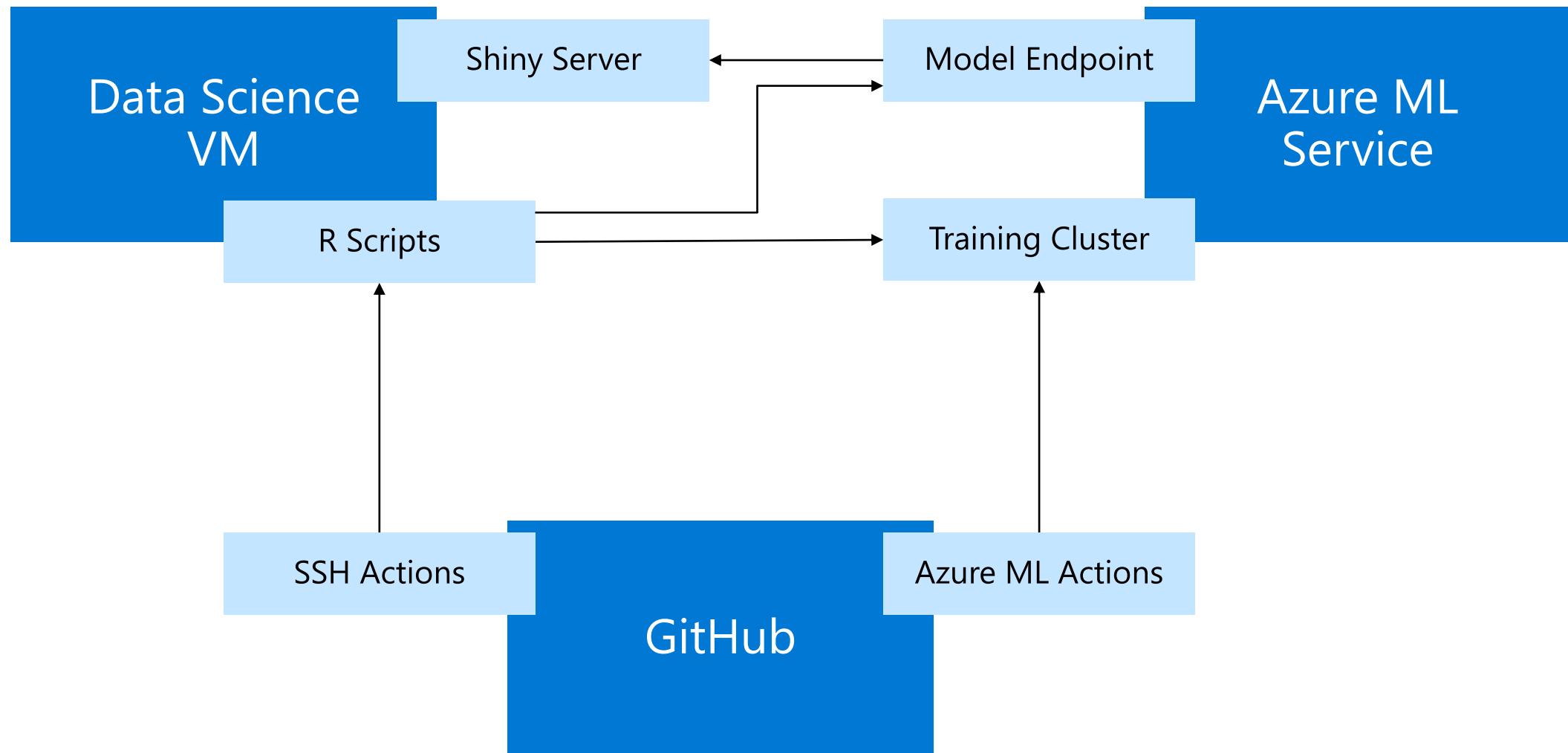
Create YAML files in `.github/workflows` to define jobs

Search GitHub Actions Marketplace for pre-defined templates

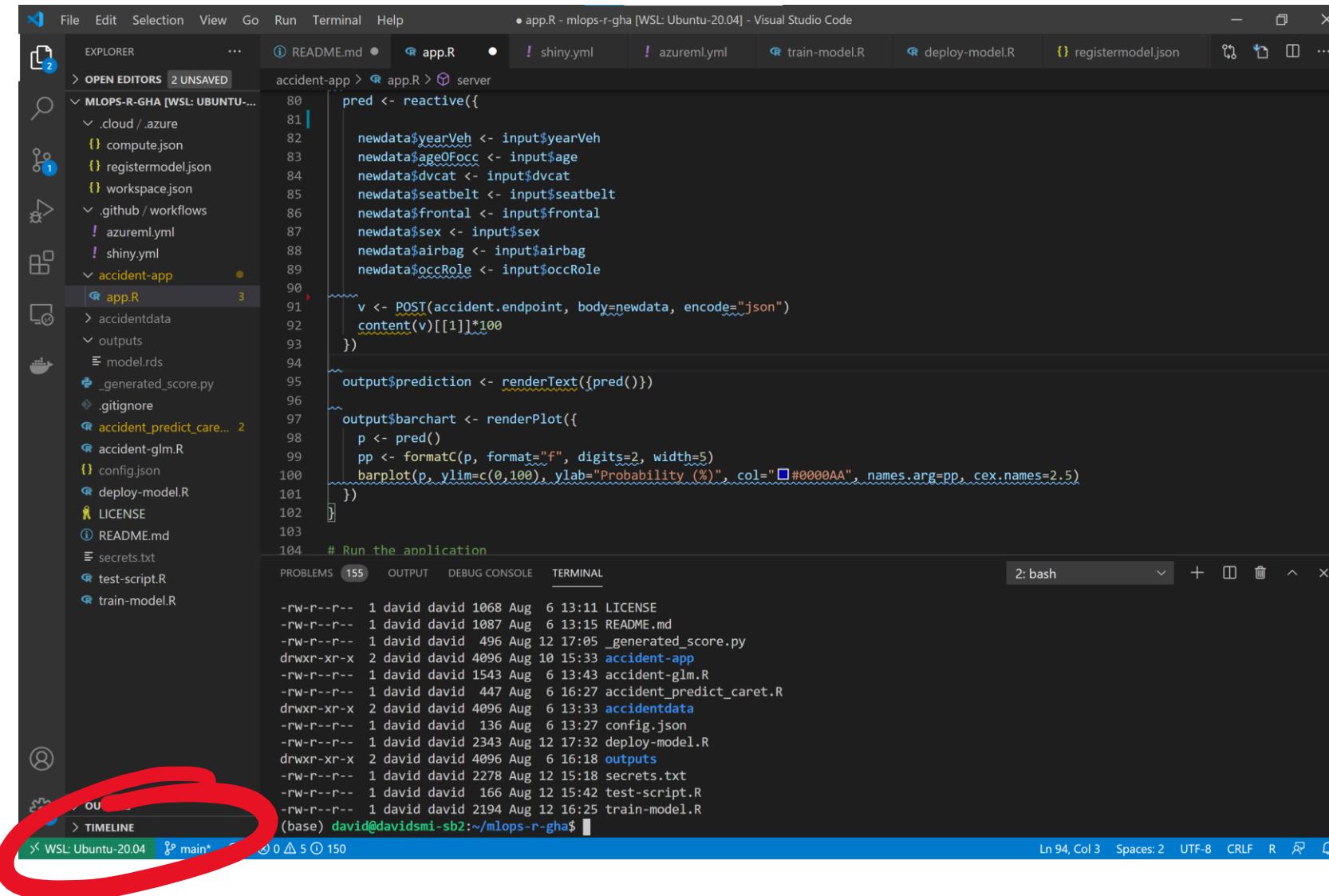
Push, then monitor workflows from the Actions tab



“Accident” Shiny Application Architecture



Dev environment: VS Code + WSL



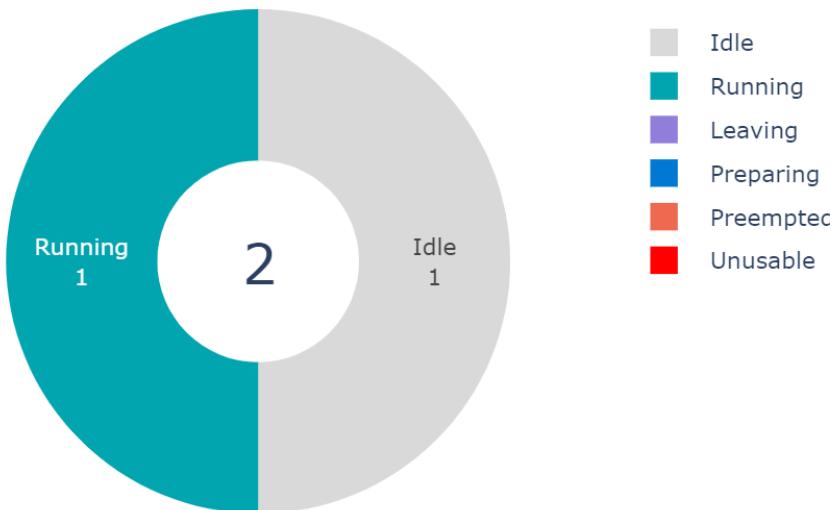
The screenshot shows the Visual Studio Code interface with the following details:

- File Explorer:** Shows the project structure under "MLOPS-R-GHA [WSL: UBUNTU-20.04]".
- Editor:** Displays an R script named "app.R". The code performs a POST request to an endpoint, processes the response, and generates a bar chart.
- Terminal:** Shows the command "david@davidsmi-sb2:~/mlops-r-gha\$".
- Status Bar:** At the bottom, it indicates the environment as "WSL: Ubuntu-20.04".

Create 4-node training cluster

azureml.yml

```
# Connect or Create a Compute Target in Azure Machine Learning
- name: Connect/Create Azure Machine Learning Compute Target
  id: aml_compute_training
  uses: Azure/aml-compute@v1
  with:
    azure_credentials: ${{ secrets.AZURE_CREDENTIALS }}
```



compute.json

```
{
  "name": "rcluster",
  "compute_type": "amlcluster",
  "min_nodes": 0,
  "max_nodes": 4,
  "idle_seconds_before_scaledown": 600
}
```

- Unused nodes de-allocate automatically
- Increase idle timeout when developing

Train model

azureml.yml

```
# Connect to the Shiny VM to train the model
- name: Train model
  uses: JimCronqvist/action-ssh@master
  env:
    AZURE_CREDENTIALS: '${{ secrets.AZURE_CREDENTIALS }}'
  with:
    hosts: ${{ secrets.SHINYUSERNAME }}@${{ secrets.SHINYHOST }}
    privateKey: ${{ secrets.SHINYKEY }}
    command: |
      cd mlops-r-gha
      export AZURE_CREDENTIALS
      Rscript train-model.R
```

train-model.R

```
est <- estimator(source_directory = ".",
                  entry_script = "accident-glm.R",
                  script_params = list("--data_folder" = ds$path(target_path)),
                  compute_target = compute_target)
run <- submit_experiment(exp, est)
```

Experiments tracked with source scripts and recorded metrics

Control execution with command line parameters

Most packages pre-loaded, custom packages supported

accident

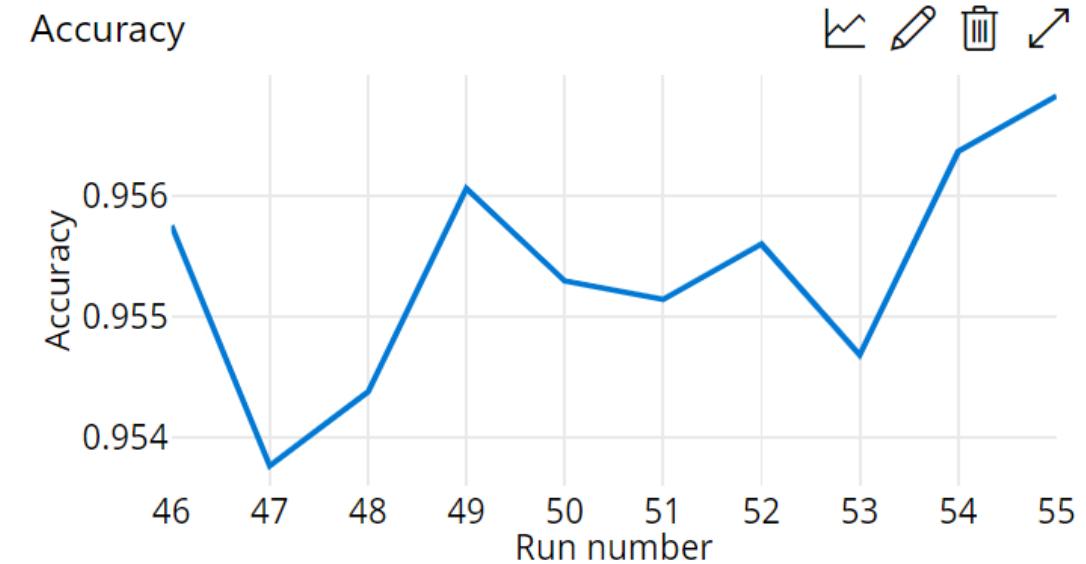
 Edit table  Refresh  Reset view  Add chart

 Add filter |  Include child runs

Run status



Accuracy



 Show only selected rows (10 selected )

	Run	Run ID	Status	Submitted time	Duration	Submitted by	Compute target
	Run 55	accide...	Completed	Aug 13, 2020 3:52 PM	4m 40s	Service principal	rcluster

Deploy Model as REST endpoint

deploy-model.R

```
inference_config <- inference_config(  
  entry_script = "accident_predict_caret.R",  
  source_directory = ".",  
  environment = r_env)  
  
aci_config <- aci_webservice_deployment_config(cpu_cores = 1, memory_gb = 0.5)  
  
aci_service <- deploy_model(ws,  
  'accidents-gha',  
  list(model),  
  inference_config,  
  aci_config)  
wait_for_deployment(aci_service, show_output = TRUE)
```

Name	Description	Created on	Created by	Updated on
accidents-gha	--	August 13, 2020 4:01 PM	d2fbafa0-336d-4ae2-83da-7c7df...	August 13, 2020 4:01 PM

Integrate model into Shiny app

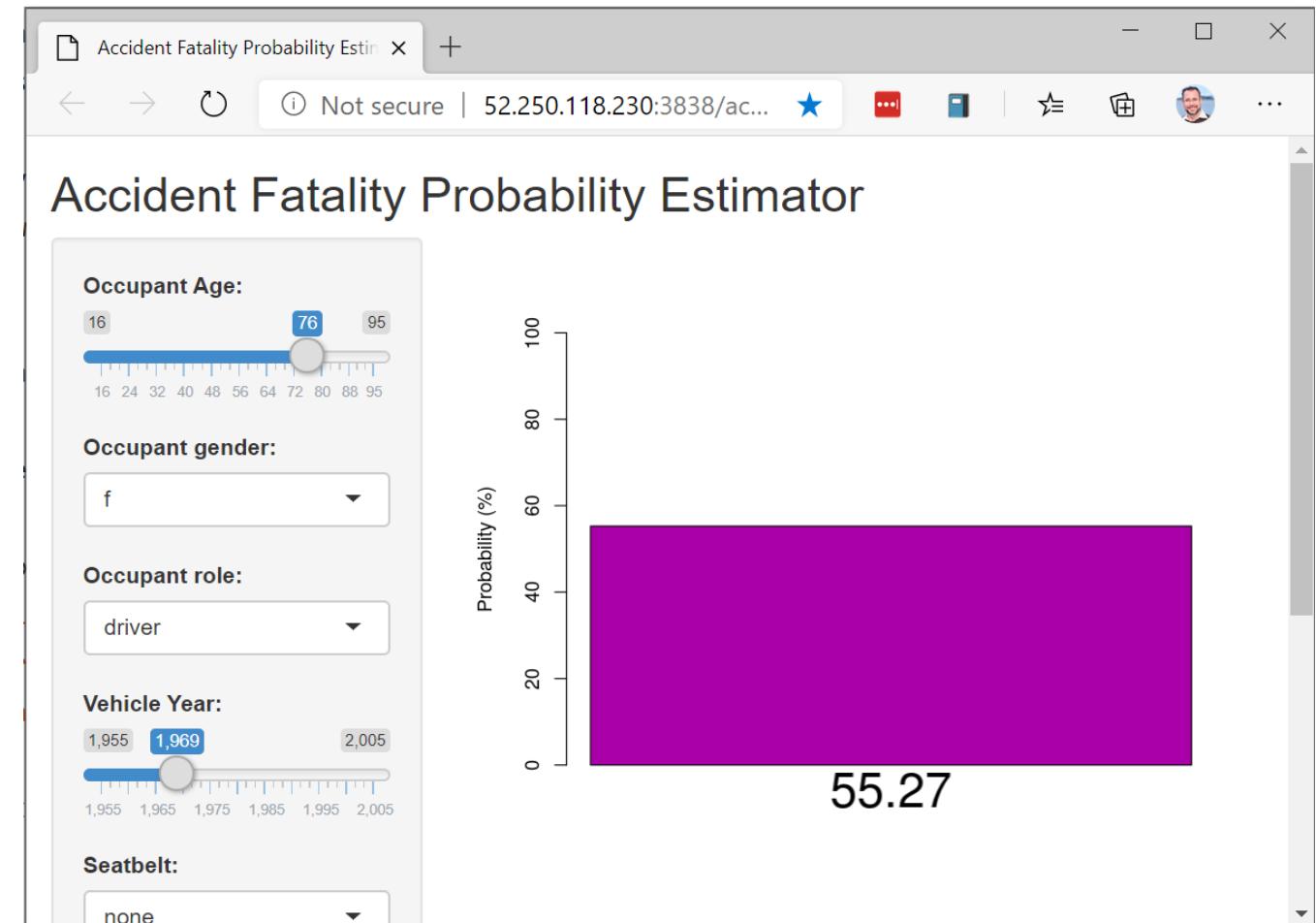
app.R

```
library(httr)
v <- POST(accident.endpoint,
           body=input,
           encode="json")
```

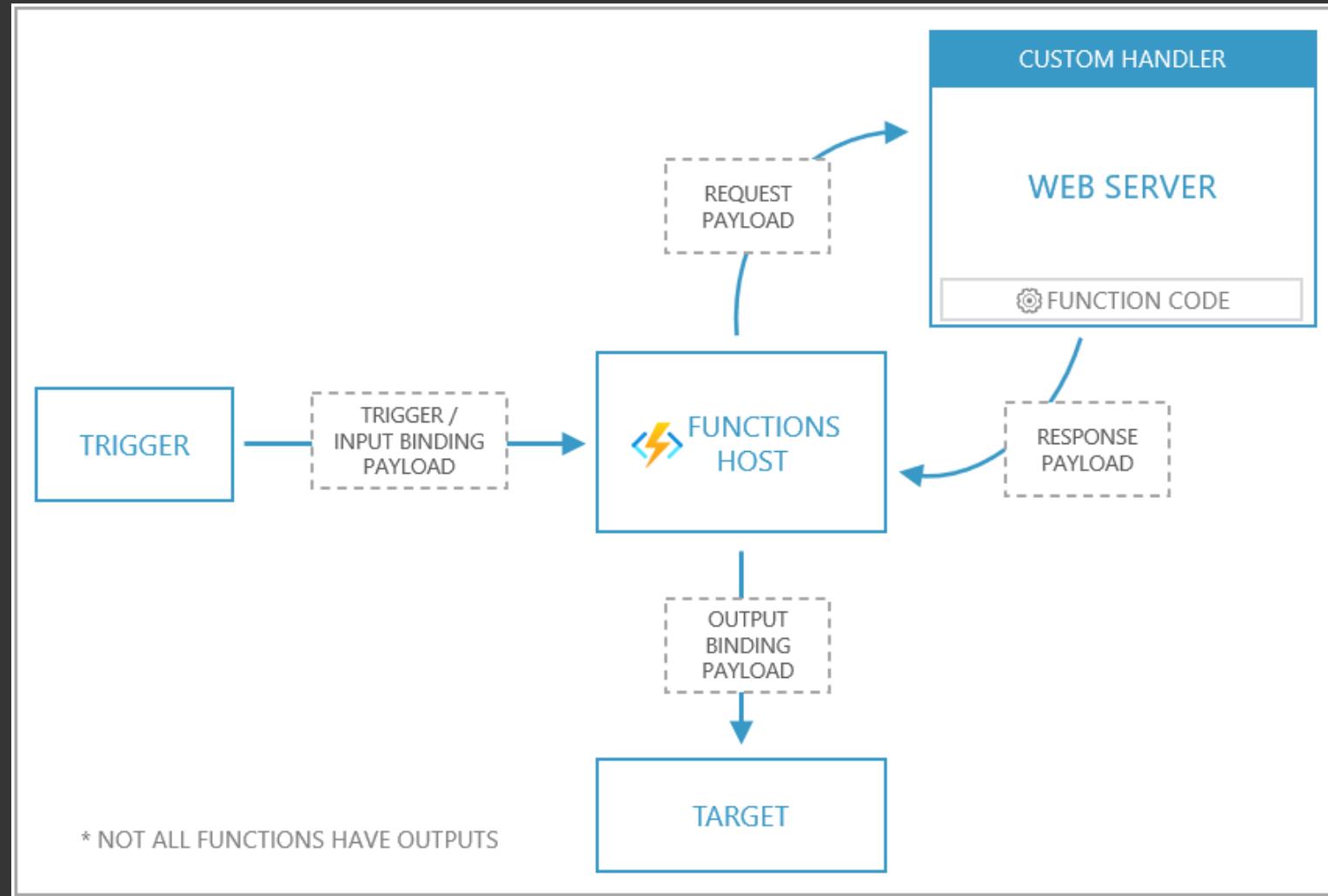
```
pred <- content(v)[[1]]
```

shiny.yml

```
- name: copy files via ssh key
  uses: appleboy/scp-action@master
  with:
    host: ${{ secrets.SHINYHOST }}
    username: ${{ secrets.SHINYUSERNAME }}
    key: ${{ secrets.SHINYKEY }}
    source: "accident-app/app.R"
    target: "~"
```



Azure Functions: Serverless R



Azure Functions Custom Handlers for R

- Create a container with R, packages, and other deps
- Create “handler” in R to implement Functions for triggers
- Push image to Docker Hub
- Use serverless endpoints in apps
- Push updated container image to update Functions

Tutorial: aka.ms/custom-handler-R

What about the future?



Thank you!

Slides, links and credits:
github.com/revodavid/R-at-Microsoft

David Smith

davidsmi@microsoft.com

@revodavid

