DSS Prototype Analysis

Alvin Murphy

Table of contents

1 DSS Prototype Overview

1.1 Installation as a Docker Container

https://github.com/jupyter/docker-stacks
https://hub.docker.com/r/jupyter/r-notebook/tags/

(optional) docker pull jupyter/r-notebook:latest

We want the Jupyter container to mount the DDS Prototype ~/analysis/ directory to provide access to scripts and data. Use the following to mount the analysis directory (i.e. current working directory) as a volume in the Juypter container. Note that the directory needed to be added as a valid mount point via the Docker Desktop Dashboard on Mac.

docker run -it –rm -d -p 10000:8888 -v ${PWD}:/home/jovyan/work –name notebook jupyter/r-notebook:latest

To find the token from the container:
docker exec -it notebook jupyter server list
or
docker logs notebook

Navigate to the container UI and enter the token: http://localhost:10000

1.2 System Context

Figure 1 depicts the context for the DSS. The DSS operator interacts with the DSS Prototype for decision assitance. The DSS relies on a aircraft database to gather real-time flight data to review in decision support algorithms.

1.3 Container Architecture

Nine containers are instantiated as part of the DSS architecuture (see Figure 2). Six provide the DSS implementation while the additional 3 support collection and calculation of metrics. Each application container was designed around the 12-Factor Application "Single Responsibility Principle"; e.g. each app has one purpose to enable rapid insertion of new capabilities with low cohesion to other functionality. At this time, all responses are canned without underlying calculations to focus on meeting the 500 ms hypothesis pryor to burdening the application with calculation latency.

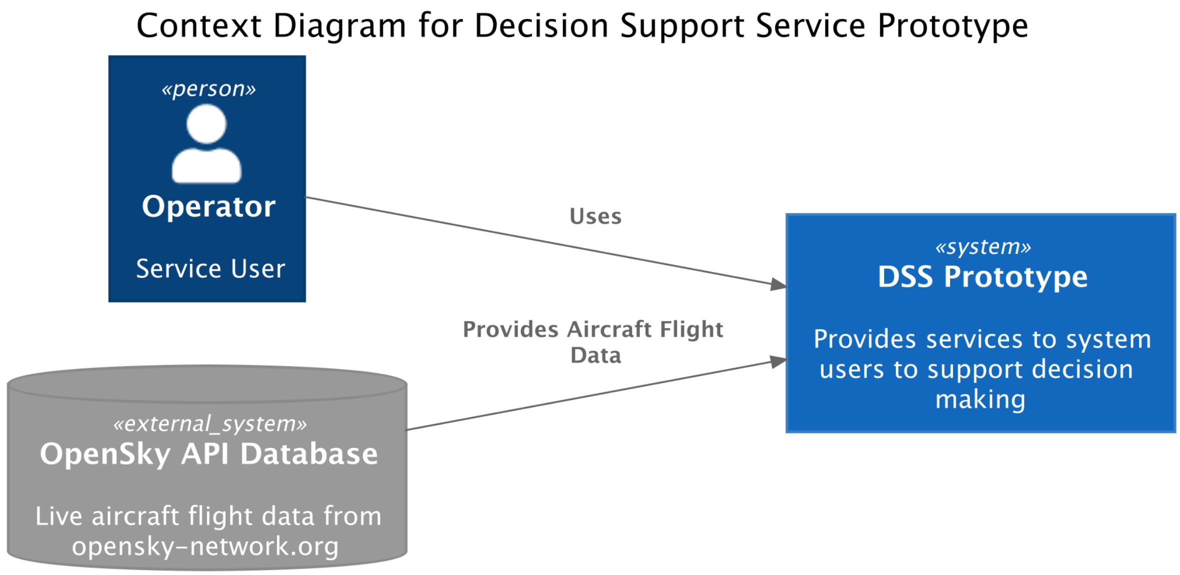


Figure 1: DSS Context Diagram

1.3.1 Applications

- opensky-int: Provides the OpenSky API for flight data. The app provides data about aircraft within 60 NM of Richmond (RIC) or Dulles (IAD) airports.
- tm-server: Provides sensor track data (e.g. OpenSky) and system tracks to support DSS services. System tracks represent the system-wide common understanding of track object states used for decision support.
- wa-app: The Weapon Assessment Application determines which weapons are capable to successfully engage a target. The wa-app uses the tm-server api to get track data.
- te-app: The Trail Engage Application predicts the success probability of an engagement with a specific weapon target pairing. The predicted track kinematic data at engagement time is provided; therefore, the current track kinematics from the tm-server are not queried prior to providing a response.
- test-app: Provides an ability to initiate automated tests. the test-app uses the dss-ui to call dss-ui endpoint to replicate operator interactions with the DSS Prototype.
- dss-ui: Provides a simple graphical interface to launch DSS services.

1.3.2 Tools

- telem-jaeger: The open source Jaeger containter collects "span" data from the DSS applications. Spans collect duration data for service calls amongst containers; e.g. latency. This the fundamental data that is being analysed here.

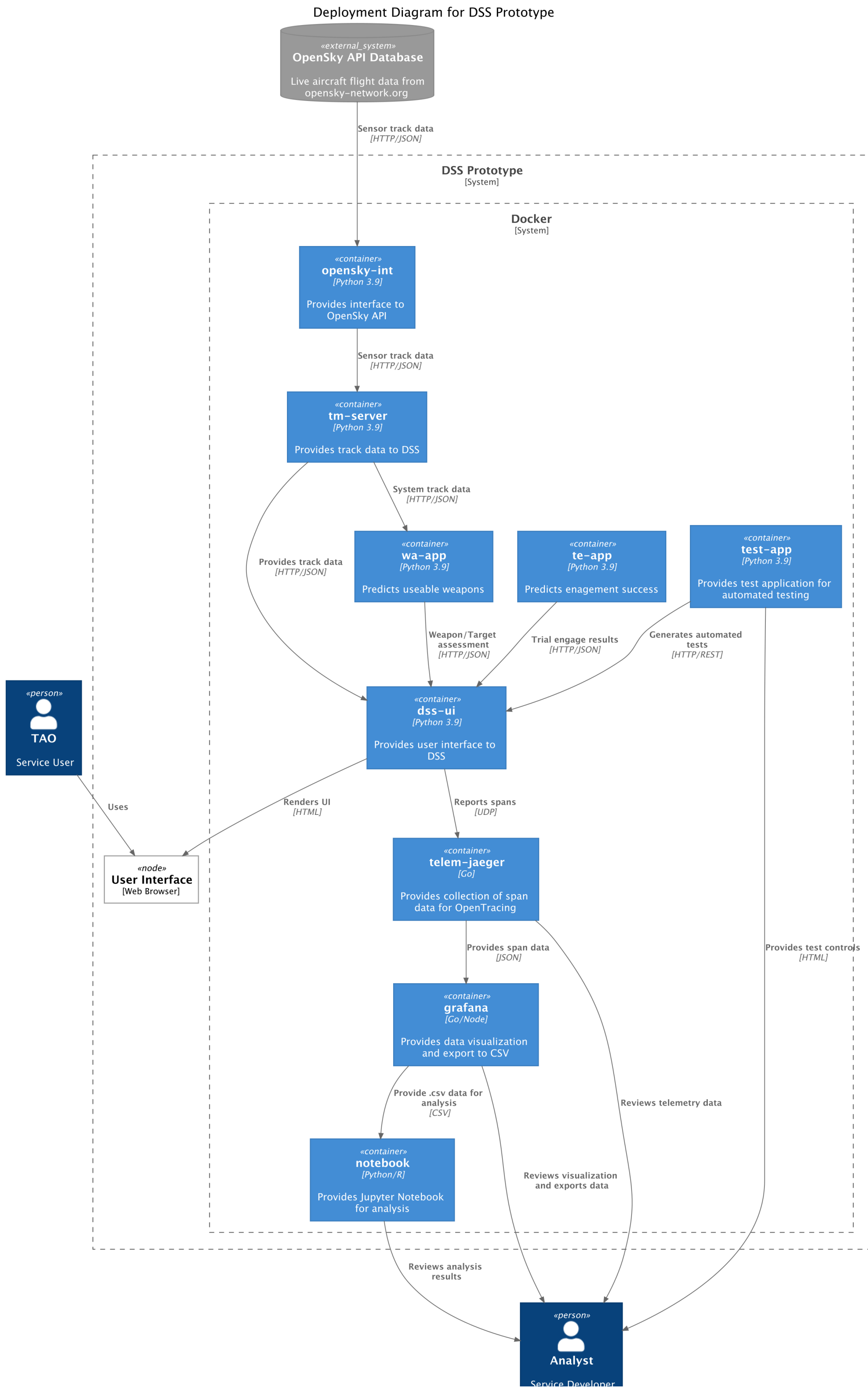


Figure 2: DSS Deployment Diagram

- grafana: The open source Grafana container connects to the telem-jaeger container to create visualization dashboards. Also, Grafana faciliates the export of data as a .csv file for analysis.
- notebook: The Jupyter Notebook container supports analysis of the data recorded by Jaeger and exported by Grafana. An embedded R software library is used for analysis.

1.4 Hypothesis

Hypotheses are "innocent until proven guilty." We'll assume that SpaceX and others have proven that DevSecOps tech can meet hard-real-time requirements but nothing available in the body of knowledge documents this.

Hypothesis: Modern DevSecOps architectures can be designed to meet hard-real-time latency (μ) requirements using modern computing environments and computing infrastructure.

$H_0 : \mu \leq 500ms$ with jitter within latency bounds
$H_a : \mu > 500ms$ with jitter exceeding latency bounds

Murphy, Alvin C. and Moreland Jr, James D. 'Integrating AI Microservices into Hard-Real-Time SoS to Ensure Trustworthiness of Digital Enterprise Using Mission Engineering'. 1 Jan. 2021 : 38 – 54.

Scrucca L., Fop M., Murphy T. B. and Raftery A. E. (2016) mclust 5: clustering, classification and density estimation using Gaussian finite mixture models The R Journal 8/1, pp. 289-317

2 Load Data Files

2.1 Review and Tag MacBook Air (2017) Data

```
   Trace.ID          Trace.name         Start.time          Duration
 Length:100         Length:100         Length:100         Length:100
 Class :character   Class :character   Class :character   Class :character
 Mode  :character   Mode  :character   Mode  :character   Mode  :character
```

A data.frame: 6 × 2

	Trace.ID <chr>	Trace.name <chr>
1	9ee3577fb1b427bc4fc17fecc5154d7d	dss-prototype: /TE
2	f05ddc4dc13aff5c3098011b2a402401	dss-prototype: /tracks
3	2bd901fbbfc9ee8dfa7c9629d93a1567	dss-prototype: /IAD
4	69a48381a14e79da08aaa2353f7db4b2	dss-prototype: /RIC
5	e83037dcb9438c04dc12fba373b5502f	dss-prototype: /WA

	Trace.ID <chr>	Trace.name <chr>
6	7e381cd880adb670bb9627ca47020938	dss-prototype: /TE

A data.frame: 6 × 2

	Start.time <chr>	Duration <chr>
1	2022-05-02 10:25:01.366	36.0 ms
2	2022-05-02 10:25:00.309	43.3 ms
3	2022-05-02 10:24:58.818	464 ms
4	2022-05-02 10:24:57.307	494 ms
5	2022-05-02 10:24:56.128	139 ms
6	2022-05-02 10:24:55.081	30.3 ms

2.1.1 Add Source Indicator to MacBook Data

2.2 Tag Linux PC (2012) Data

2.3 Tag Raspberry Pi 4 (2020) Data

2.4 Tag AWS EC2 t2.micro Data

2.5 Tag ODU CCI Data

2.6 Merge Data Files

A data.frame: 500 × 6

Trace.ID <chr>	Trace.name <chr>	Start.time <chr>	Duration <chr>	platform <chr>	env <dbl>
9ee3577fb1b427bc4fc17fecc5154d7d	dss-prototype: /TE	2022-05-02 10:25:01.366	36.0 ms	2017-macbook	0
f05ddc4dc13aff5d3098011b2a402401	dss-prototype: /tracks	2022-05-02 10:25:00.309	43.3 ms	2017-macbook	0
2bd901fbbfc9ee8dfa7c9629d93a1567	dss-prototype: /IAD	2022-05-02 10:24:58.818	464 ms	2017-macbook	0

Trace.ID <chr>	Trace.name <chr>	Start.time <chr>	Duration <chr>	platform <chr>	env <dbl>
69a48381a14e79d08aaa2353f7db4b2	das-prototype: /RIC	2022-05-02 10:24:57.307	494 ms	2017-macbook	0
e83037dcb9438cd4dc12fba373b5502f	das-prototype: /WA	2022-05-02 10:24:56.128	139 ms	2017-macbook	0
7e381cd880adb670bb9627ca47020938	das-prototype: /TE	2022-05-02 10:24:55.081	30.3 ms	2017-macbook	0
092e01448c8f39b99d39c60c456cda72	das-prototype: /tracks	2022-05-02 10:24:54.040	30.0 ms	2017-macbook	0
55f2710ea10d84cdf8ba9e5bf31ce719	das-prototype: /IAD	2022-05-02 10:24:52.545	478 ms	2017-macbook	0
d1a0499b11129a50393aaa1f6e478df2	das-prototype: /RIC	2022-05-02 10:24:50.974	546 ms	2017-macbook	0
68208a03967e73d1abdd626096ab3525	das-prototype: /WA	2022-05-02 10:24:49.891	70.7 ms	2017-macbook	0
0379e864afb13ed92e09235c8715f1a2	das-prototype: /TE	2022-05-02 10:24:48.849	24.5 ms	2017-macbook	0
002df2c1fe34daa6c39ceb3cb6d20072	das-prototype: /tracks	2022-05-02 10:24:47.706	126 ms	2017-macbook	0
2fdb400d91125d68ecbb0a6416a22952	das-prototype: /IAD	2022-05-02 10:24:46.168	398 ms	2017-macbook	0
c03154352b55d78a2ca57cc9bf74c322	das-prototype: /RIC	2022-05-02 10:24:44.714	442 ms	2017-macbook	0
862e3e7d784e40c9aeb94ea7b53212df	das-prototype: /WA	2022-05-02 10:24:43.625	74.8 ms	2017-macbook	0
ea6c6e6f09eea12091518e238213b302	das-prototype: /TE	2022-05-02 10:24:42.562	36.5 ms	2017-macbook	0

Trace.ID <chr>	Trace.name <chr>	Start.time <chr>	Duration <chr>	platform <chr>	env <dbl>
8c363c4044084c32e64cb04281f4825a	dss-prototype: /tracks	2022-05-02 10:24:41.527	20.5 ms	2017-macbook	0
f7997247087499b61180a5d1002568ce	dss-prototype: /IAD	2022-05-02 10:24:40.004	510 ms	2017-macbook	0
01a08abf31a281816c24a67222b181e5	dss-prototype: /RIC	2022-05-02 10:24:38.400	579 ms	2017-macbook	0
02d67201ea1c144b8c4419a23a15b0dd	dss-prototype: /WA	2022-05-02 10:24:37.001	369 ms	2017-macbook	0
5e64e301a7dd6ddb25c1ce70063d4c74	dss-prototype: /TE	2022-05-02 10:24:35.948	30.1 ms	2017-macbook	0
8b14791720bc81d538a092b431c602bd	dss-prototype: /tracks	2022-05-02 10:24:34.903	27.6 ms	2017-macbook	0
0140e81c442c31744bea4bff26a98c01	dss-prototype: /IAD	2022-05-02 10:24:33.487	403 ms	2017-macbook	0
b15c1e3efb6650824b63a6f935638314	dss-prototype: /RIC	2022-05-02 10:24:32.064	410 ms	2017-macbook	0
8bb292584b55356be7772855432f27f7	dss-prototype: /WA	2022-05-02 10:24:30.969	80.0 ms	2017-macbook	0
d153b3205003467d08b6816f62d8e1a1	dss-prototype: /TE	2022-05-02 10:24:29.906	46.7 ms	2017-macbook	0
15f7e65d2d84a36d4455179dfb849b3d	dss-prototype: /tracks	2022-05-02 10:24:28.865	15.5 ms	2017-macbook	0
a5b0d08991c9073f93dcb02639deceed	dss-prototype: /IAD	2022-05-02 10:24:27.220	632 ms	2017-macbook	0
e113d8c671024bd2a15245cff74338e1	dss-prototype: /RIC	2022-05-02 10:24:25.534	669 ms	2017-macbook	0

Trace.ID <chr>	Trace.name <chr>	Start.time <chr>	Duration <chr>	platform <chr>	env <dbl>
db45212184a7c9d0d9b8c7c21a352dea	dss-prototype: /WA	2022-05-02 10:24:24.476	38.6 ms	2017-macbook	0
0ed9903f816c28ed2327b20d7ccb7ed4	dss-prototype: /TE	2022-06-28 21:17:13.236	9.04 ms	2022-odu-cci	4
c093ae490db58e71c13958aa69946563	dss-prototype: /tracks	2022-06-28 21:17:12.226	4.77 ms	2022-odu-cci	4
1ce5a0e853ee2bbbf73322e696b82cd2	dss-prototype: /IAD	2022-06-28 21:17:10.731	490 ms	2022-odu-cci	4
708c66352a157527facdd200d4a34012	dss-prototype: /RIC	2022-06-28 21:17:09.280	446 ms	2022-odu-cci	4
997cd2170b7dfb09ad0929c43a4b6072	dss-prototype: /WA	2022-06-28 21:17:08.263	11.9 ms	2022-odu-cci	4
da07cdf269403cd74dd3193d4c5c0a82	dss-prototype: /TE	2022-06-28 21:17:07.251	6.84 ms	2022-odu-cci	4
dcc9a36a1b37e63442353c9e0e8096f2	dss-prototype: /tracks	2022-06-28 21:17:06.241	4.96 ms	2022-odu-cci	4
d16b9bdfb9cba5db2b488df2318fc499	dss-prototype: /IAD	2022-06-28 21:17:04.923	313 ms	2022-odu-cci	4
347cdd652125000438 8e872532dd8821	dss-prototype: /RIC	2022-06-28 21:17:03.408	509 ms	2022-odu-cci	4
d8e3417f95b02fa197709c9e8317b0e2	dss-prototype: /WA	2022-06-28 21:17:02.392	11.0 ms	2022-odu-cci	4
5e2b1a72df41ca1b4b21074df6965042	dss-prototype: /TE	2022-06-28 21:17:01.380	7.42 ms	2022-odu-cci	4
1aae425e48e19d4923 7485a70717af12	dss-prototype: /tracks	2022-06-28 21:17:00.370	5.11 ms	2022-odu-cci	4

Trace.ID <chr>	Trace.name <chr>	Start.time <chr>	Duration <chr>	platform <chr>	env <dbl>
3ace4d30e1fdc05[illegible]29b6cd28a310[illegible]	dsa-prototype: /IAD	2022-06-28 21:16:58.992	372 ms	2022-odu-cci	4
23dc17c1f02b5ed[illegible]3f44c3f41ecabd[illegible]	dsa-prototype: /RIC	2022-06-28 21:16:57.512	475 ms	2022-odu-cci	4
1f5d843e1210340[illegible]4662e8783a10c[illegible]	dsa-prototype: /WA	2022-06-28 21:16:56.495	12.2 ms	2022-odu-cci	4
9d4774dc3d59dc0[illegible]033130809a51[illegible]	dsa-prototype: /TE	2022-06-28 21:16:55.481	8.86 ms	2022-odu-cci	4
ada8933d70e52c8[illegible]acecf76ff853b[illegible]	dsa-prototype: /tracks	2022-06-28 21:16:54.471	5.29 ms	2022-odu-cci	4
30053467acbb832[illegible]f1e732bc718fd[illegible]	dsa-prototype: /IAD	2022-06-28 21:16:53.079	387 ms	2022-odu-cci	4
77bf5e9cce7b61e[illegible]cdd1ca743d4[illegible]	dsa-prototype: /RIC	2022-06-28 21:16:51.416	657 ms	2022-odu-cci	4
1995947229c866[illegible]9e9932a212ef[illegible]	dsa-prototype: /WA	2022-06-28 21:16:50.400	11.6 ms	2022-odu-cci	4
d5a64dbe13edf9[illegible]f5dd3383018[illegible]	dsa-prototype: /TE	2022-06-28 21:16:49.387	7.87 ms	2022-odu-cci	4
b539b6eb5ff3d4e[illegible]867579c5259d[illegible]	dsa-prototype: /tracks	2022-06-28 21:16:48.376	5.93 ms	2022-odu-cci	4
3c5fb1a8e2ff6bb[illegible]9b69fb75fd52[illegible]	dsa-prototype: /IAD	2022-06-28 21:16:47.007	364 ms	2022-odu-cci	4
e33ce1f66630c58[illegible]b027ad73eb[illegible]	dsa-prototype: /RIC	2022-06-28 21:16:45.681	321 ms	2022-odu-cci	4
8c414573b6a50f6[illegible]ba9b8f1e7942[illegible]	dsa-prototype: /WA	2022-06-28 21:16:44.664	11.1 ms	2022-odu-cci	4

Trace.ID <chr>	Trace.name <chr>	Start.time <chr>	Duration <chr>	platform <chr>	env <dbl>
9af11db84880a023db4b37c98c967b2a	dss-prototype: /TE	2022-06-28 21:16:43.651	8.37 ms	2022-odu-cci	4
ad3b002e6777f1ed0e037de96121842	dss-prototype: /tracks	2022-06-28 21:16:42.640	5.25 ms	2022-odu-cci	4
7208428dd3d2b5d1b113f3fbb0fc35db5	dss-prototype: /IAD	2022-06-28 21:16:41.285	350 ms	2022-odu-cci	4
83425a8c1972bed340215628255762d	dss-prototype: /RIC	2022-06-28 21:16:39.728	551 ms	2022-odu-cci	4
8d5c3f143cdeffc8903cad84991a3af	dss-prototype: /WA	2022-06-28 21:16:38.711	11.2 ms	2022-odu-cci	4

3 Convert Data into Useable Metrics

To make the data more usable and easier to understand we apply conversions from text to numeric and add additional columns with supporting information. A **useCase** column is added to identify specific DSS request use cases; e.g. Get Dulles Airport Data. The data also indicates whether the request is managed internally or a connection to an external service is required to provided a response (i.e., https://opensky-network.org). A **numContainers** column is added to indicate the number of containers involved in providing a use case response (e.g. independent variable). An **ext** column is added to indicate whether an API external to the Docker environment is used; e.g., ext = TRUE for OpenSky API calls.

3.1 Add Additional Column Descriptors

```
   Trace.ID           Trace.name          Start.time          Duration     
 Length:500         Length:500         Min.   :1.651e+09   Min.   :   4.29  
 Class :character   Class :character   1st Qu.:1.655e+09   1st Qu.:   7.42  
 Mode  :character   Mode  :character   Median :1.655e+09   Median :  21.65  
                                       Mean   :1.654e+09   Mean   : 198.15  
                                       3rd Qu.:1.655e+09   3rd Qu.: 381.00  
                                       Max.   :1.656e+09   Max.   :2000.00  
   platform              env      useCase             useCaseNum    ext         
 Length:500         Min.   :0   Length:500         Min.   :1    Mode :logical  
```

```
 Class :character   1st Qu.:1   Class :character   1st Qu.:2   FALSE:300
 Mode  :character   Median :2   Mode  :character   Median :3   TRUE :200
                    Mean   :2                      Mean   :3
                    3rd Qu.:3                      3rd Qu.:4
                    Max.   :4                      Max.   :5
```

A data.frame: 6 × 5

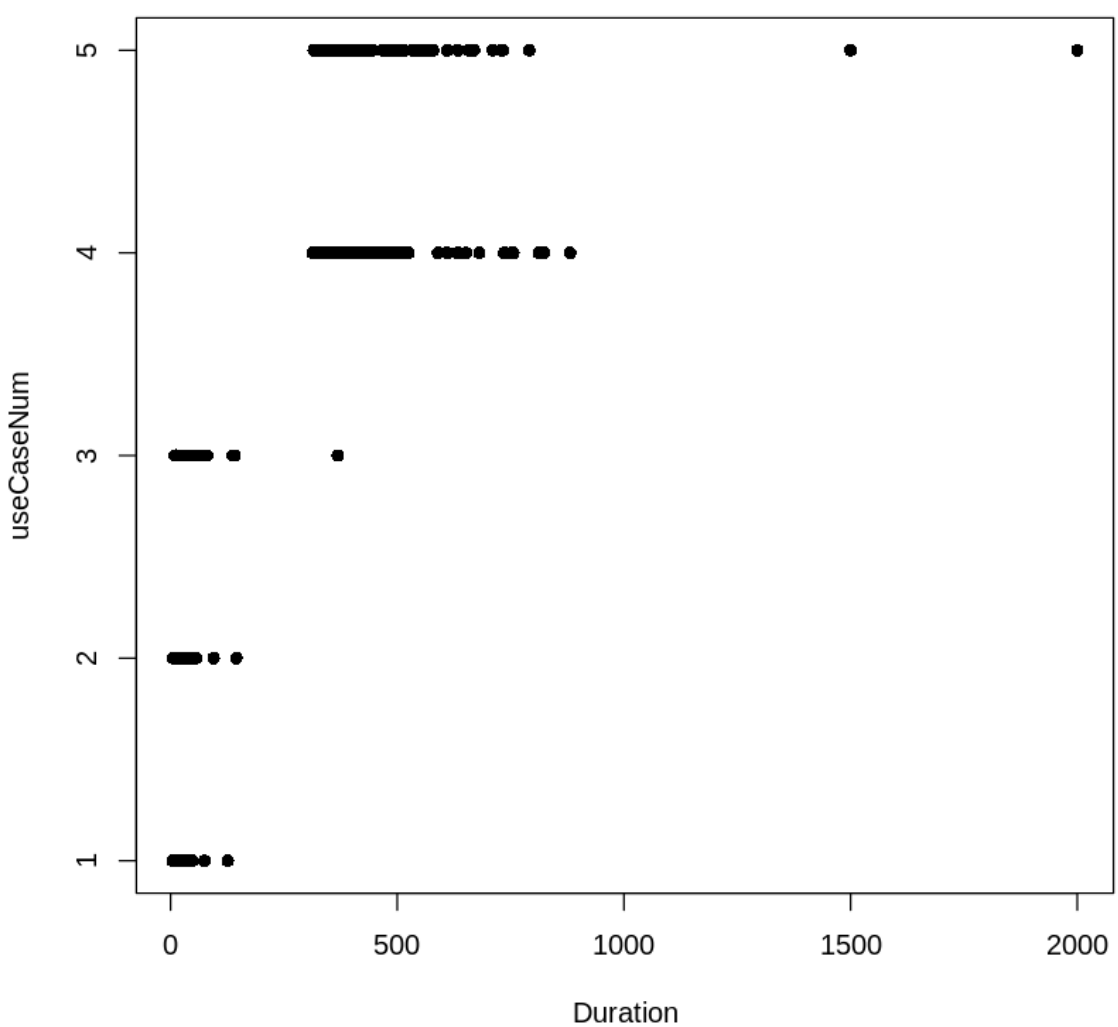
	Trace.ID <chr>	Trace.name <chr>	Start.time <dbl>	Duration <dbl>	platform <chr>
1	d2e7	/tracks	1654551325	4.29	2012-linpc
2	d2e7	/tracks	1654551325	4.29	2020-rpi4
3	813c	/tracks	1654551348	4.32	2012-linpc
4	813c	/tracks	1654551348	4.32	2020-rpi4
5	7aae	/tracks	1654551372	4.39	2012-linpc
6	7aae	/tracks	1654551372	4.39	2020-rpi4

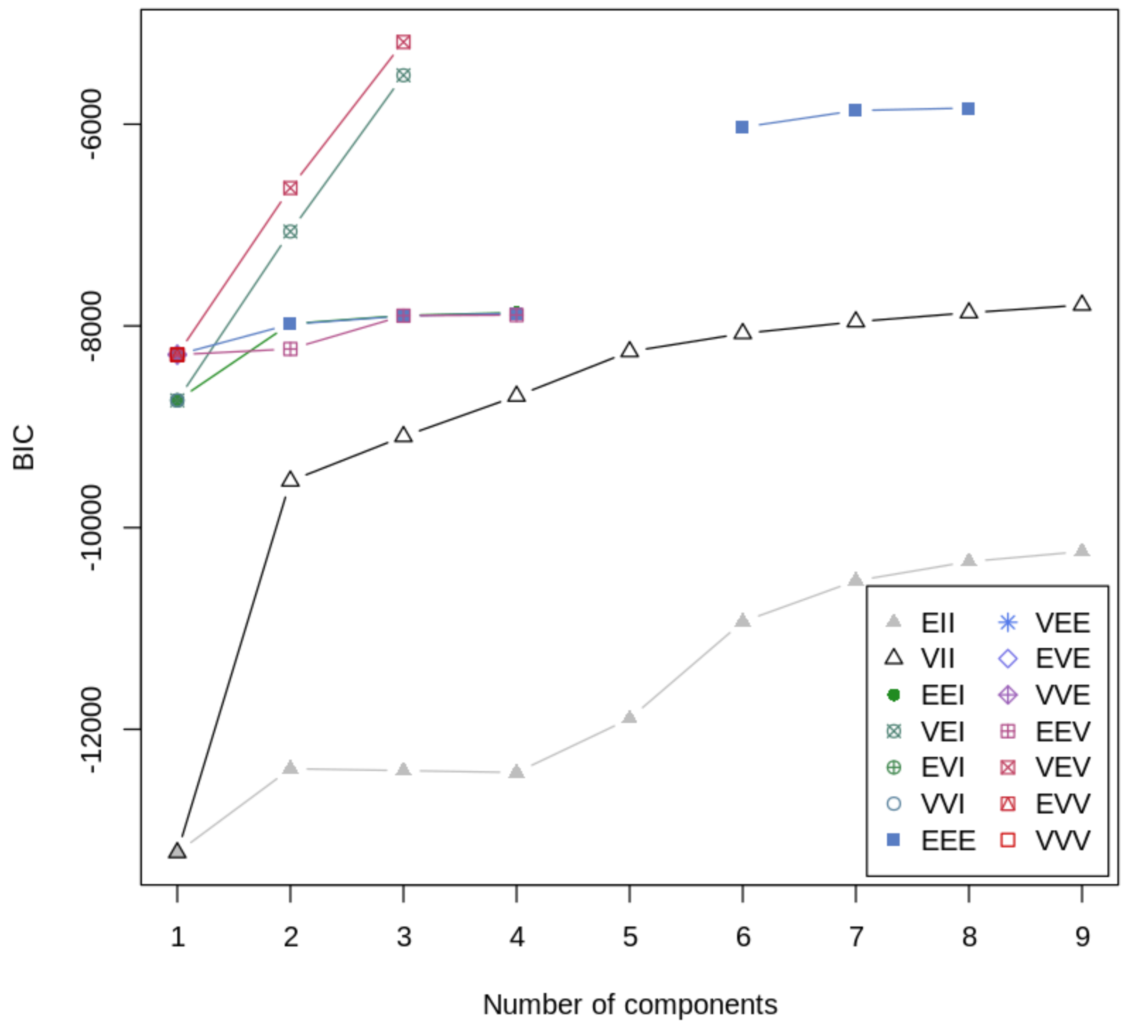
A data.frame: 6 × 4

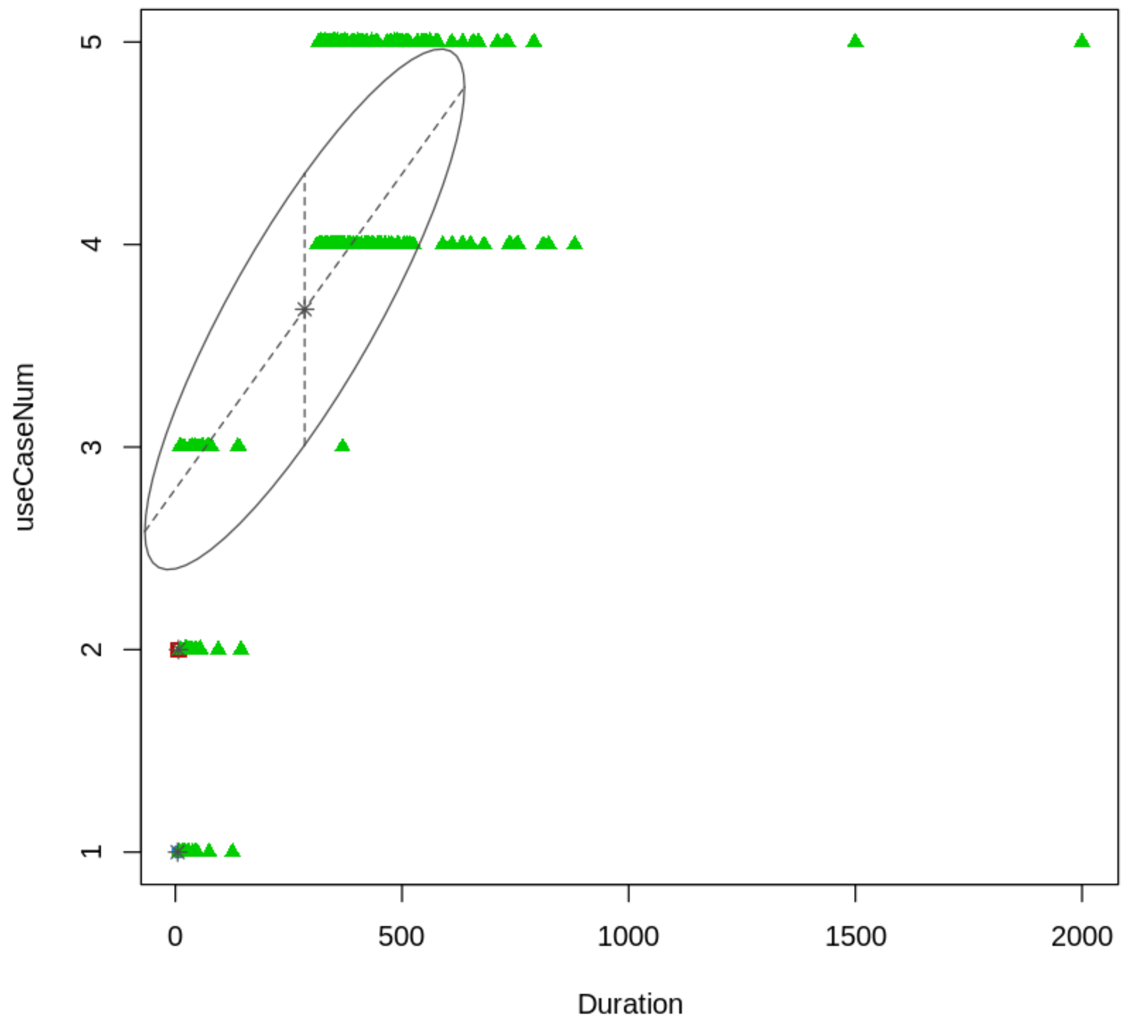
	env <dbl>	useCase <chr>	useCaseNum <dbl>	ext <lgl>
1	1	Get Stored Local DSS Tracks (Internal)	1	FALSE
2	2	Get Stored Local DSS Tracks (Internal)	1	FALSE
3	1	Get Stored Local DSS Tracks (Internal)	1	FALSE
4	2	Get Stored Local DSS Tracks (Internal)	1	FALSE
5	1	Get Stored Local DSS Tracks (Internal)	1	FALSE

	env <dbl>	useCase <chr>	useCaseNum <dbl>	ext <lgl>
6	2	Get Stored Local DSS Tracks (Internal)	1	FALSE

3.2 Exploratory Analysis Plots

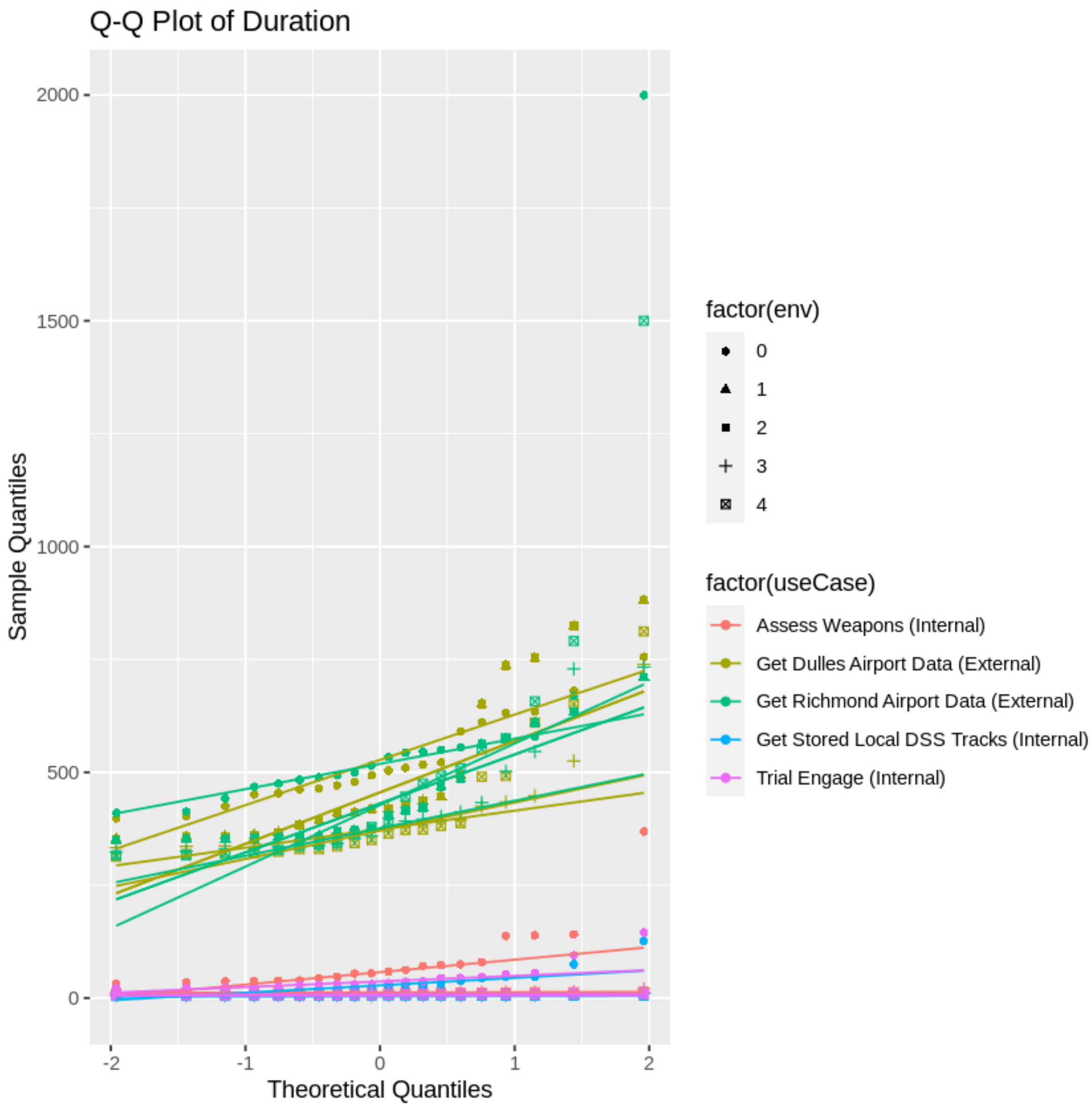



BIC
-6000
-8000
-10000
-12000
1
2
3
4
5
6
7
8
9
Number of components
EII
VII
EEI
VEI
EVI
VVI
EEE
VEE
EVE
VVE
EEV
VEV
EVV
VVV




```
'Mclust' model object: (VEV,3)

Available components:
 [1] "call"           "data"           "modelName"      "n"
 [5] "d"              "G"              "BIC"            "loglik"
 [9] "df"             "bic"            "icl"            "hypvol"
[13] "parameters"     "z"              "classification" "uncertainty"
```

3.3 Q-Q Normality Test




```
    Shapiro-Wilk normality test

data:  spanMetrics$Duration
W = 0.7464, p-value < 2.2e-16
```

A transformation is needed to apply statistical analysis.

4 Clean the Data

4.1 Search for Outliers

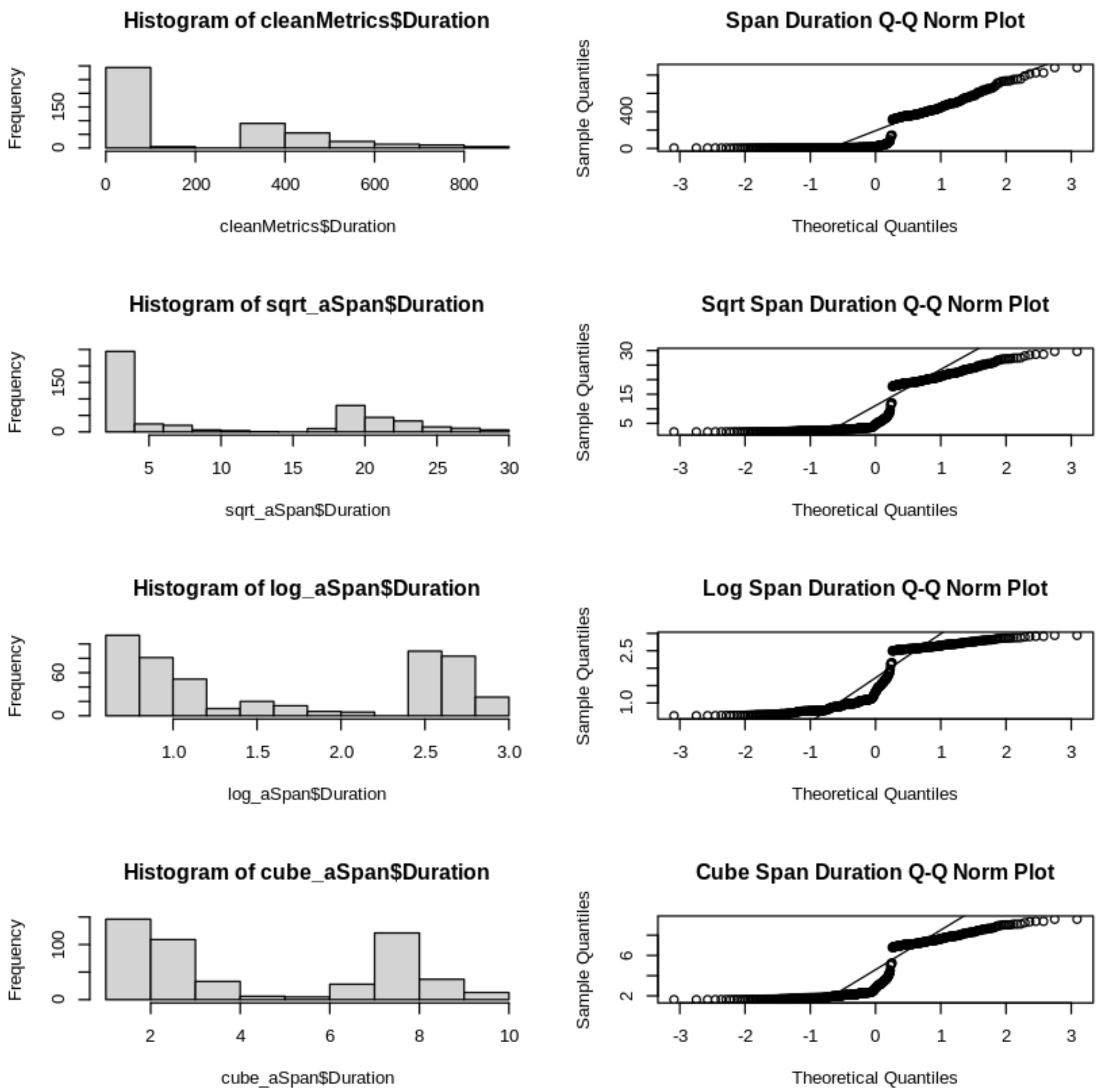
1. 1500
2. 2000

4.2 Normality Testing of Each Environment

```
    Shapiro-Wilk normality test

data:  env0MacSubset$Duration
W = 0.78089, p-value = 7.724e-11

    Shapiro-Wilk normality test

data:  env1LinSubset$Duration
W = 0.74353, p-value = 6.337e-12

    Shapiro-Wilk normality test

data:  env2PiSubset$Duration
W = 0.74353, p-value = 6.337e-12

    Shapiro-Wilk normality test

data:  env3AWSSubset$Duration
W = 0.73428, p-value = 3.67e-12

    Shapiro-Wilk normality test

data:  env4CCI_Subset$Duration
W = 0.74182, p-value = 6.676e-12
```

4.3 Transformation of Clean Metrics

4.3.1 Sqrt, Log, and Cube Transformations



None of these transformation yield distributions that would be considered normal. Most likely due to access to external and internal services with differing latency. Lets try another transformation.

Our assumption here is that the separation of **Sample Quantiles** is from the difference between internal and external span durations (e.g. latency). Let's see what happens when we split the samples.

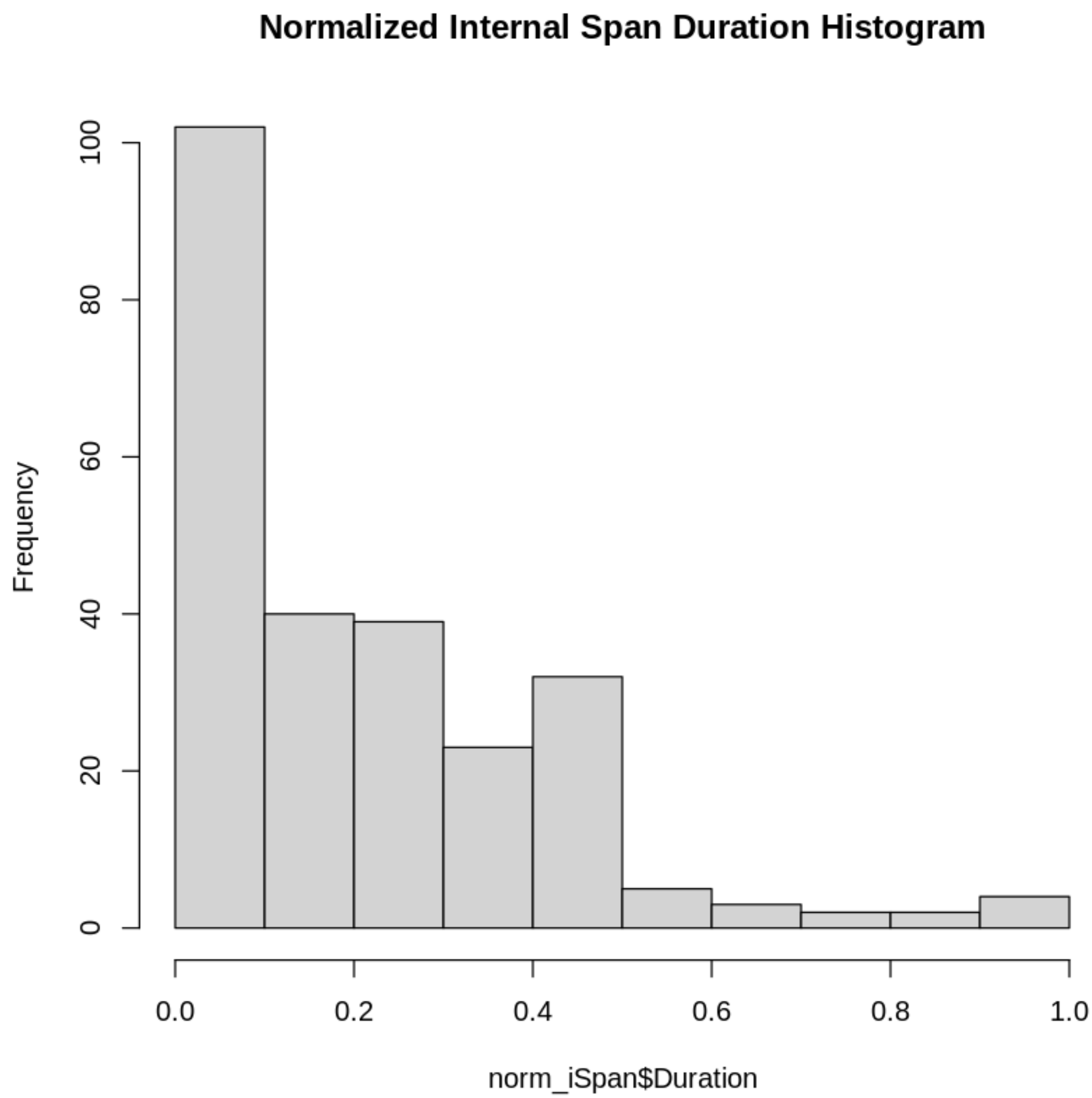
5 Separating "Clean" Internal from External Data

5.1 Internal Data

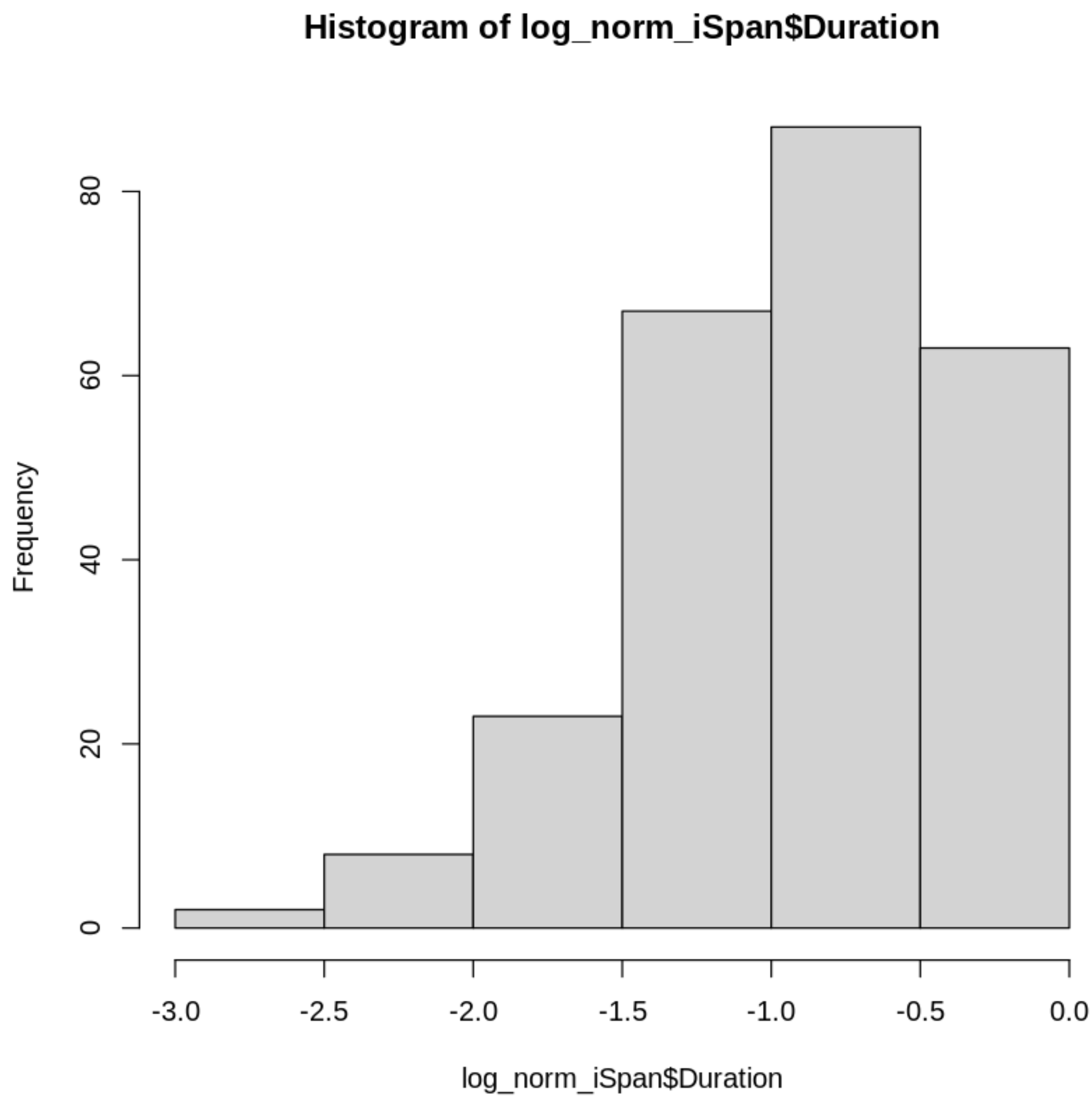
1. 27.6
2. 28.3
3. 28.5
4. 30
5. 38.1
6. 43.3
7. 46.5
8. 47.7
9. 74.6
10. 126
11. 23.9
12. 24.5
13. 26.2
14. 29.2
15. 29.4
16. 30.1
17. 30.3
18. 30.5
19. 32.9
20. 36
21. 36.5
22. 44.3
23. 44.8
24. 46.7
25. 53.1
26. 55.8
27. 94.9
28. 145
29. 32.3
30. 35.1
31. 37.5
32. 37.7
33. 38.6
34. 39

35. 43.5
36. 47.5
37. 54.3
38. 54.6
39. 58.8
40. 62.1
41. 70.7
42. 72.9
43. 74.8
44. 80
45. 137
46. 139
47. 141
48. 369

```
    Shapiro-Wilk normality test

data:  internalSpanData$Duration
W = 0.38727, p-value < 2.2e-16

    Shapiro-Wilk normality test

data:  dssSpanData$Duration
W = 0.38853, p-value < 2.2e-16

    Shapiro-Wilk normality test

data:  iSpan$Duration
W = 0.86794, p-value = 6.384e-14
```




```
    Shapiro-Wilk normality test

data:  norm_iSpan$Duration
W = 0.86794, p-value = 6.384e-14
```



5.2 External Data

1. 791
2. 753
3. 753
4. 756
5. 812
6. 824

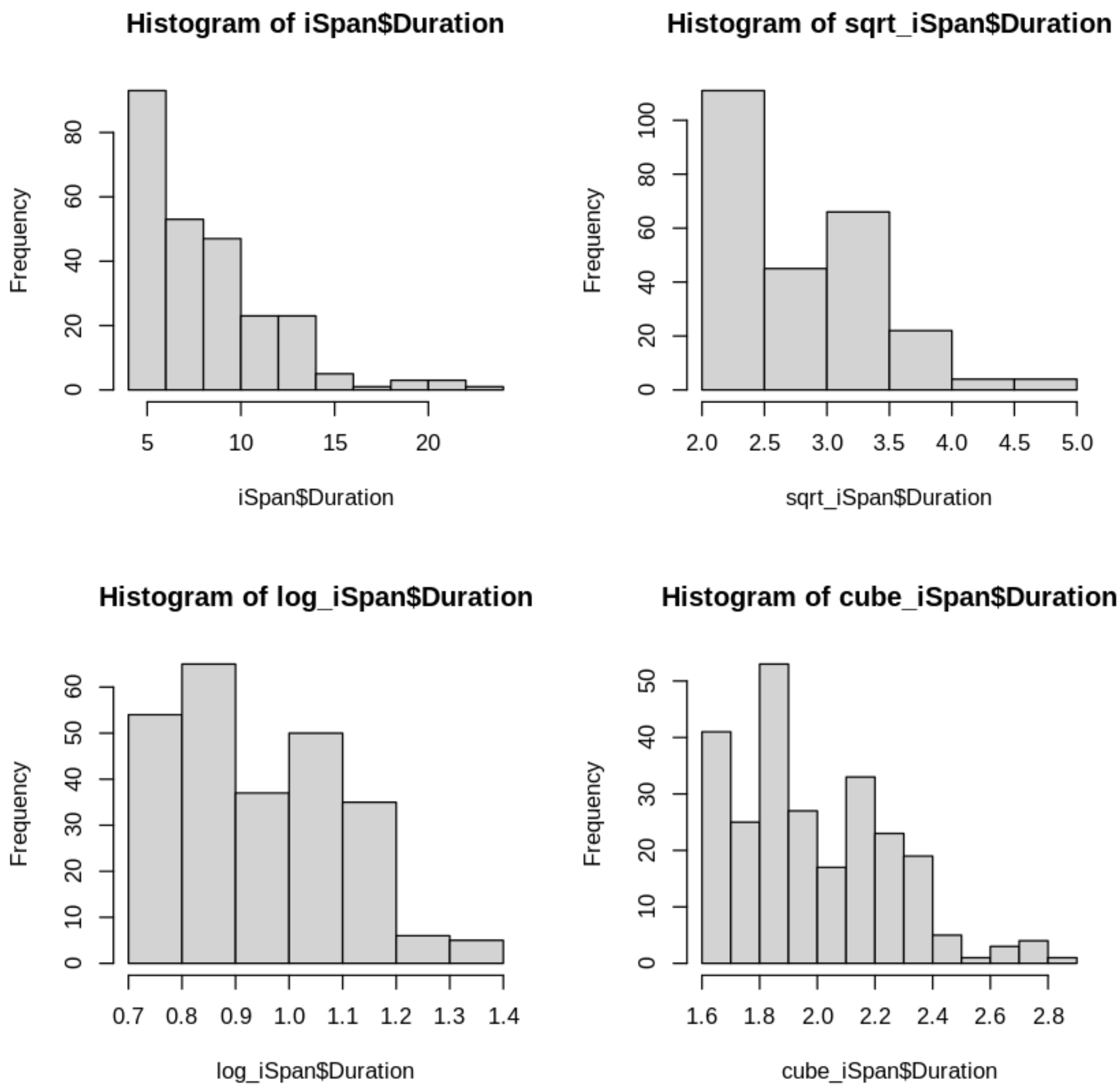
7. 824
8. 881
9. 881

```
    Shapiro-Wilk normality test

data:  externalSpanData$Duration
W = 0.85885, p-value = 1.398e-12

    Shapiro-Wilk normality test

data:  eSpan$Duration
W = 0.88201, p-value = 4.995e-11

    Shapiro-Wilk normality test

data:  env0MacE_Subset$Duration
W = 0.78089, p-value = 7.724e-11

    Shapiro-Wilk normality test

data:  env1LinE_Subset$Duration
W = 0.74353, p-value = 6.337e-12

    Shapiro-Wilk normality test

data:  env2PiE_Subset$Duration
W = 0.74353, p-value = 6.337e-12

    Shapiro-Wilk normality test

data:  env3AWS_E_Subset$Duration
W = 0.73428, p-value = 3.67e-12

    Shapiro-Wilk normality test

data:  env4CCI_E_Subset$Duration
W = 0.74182, p-value = 6.676e-12
```

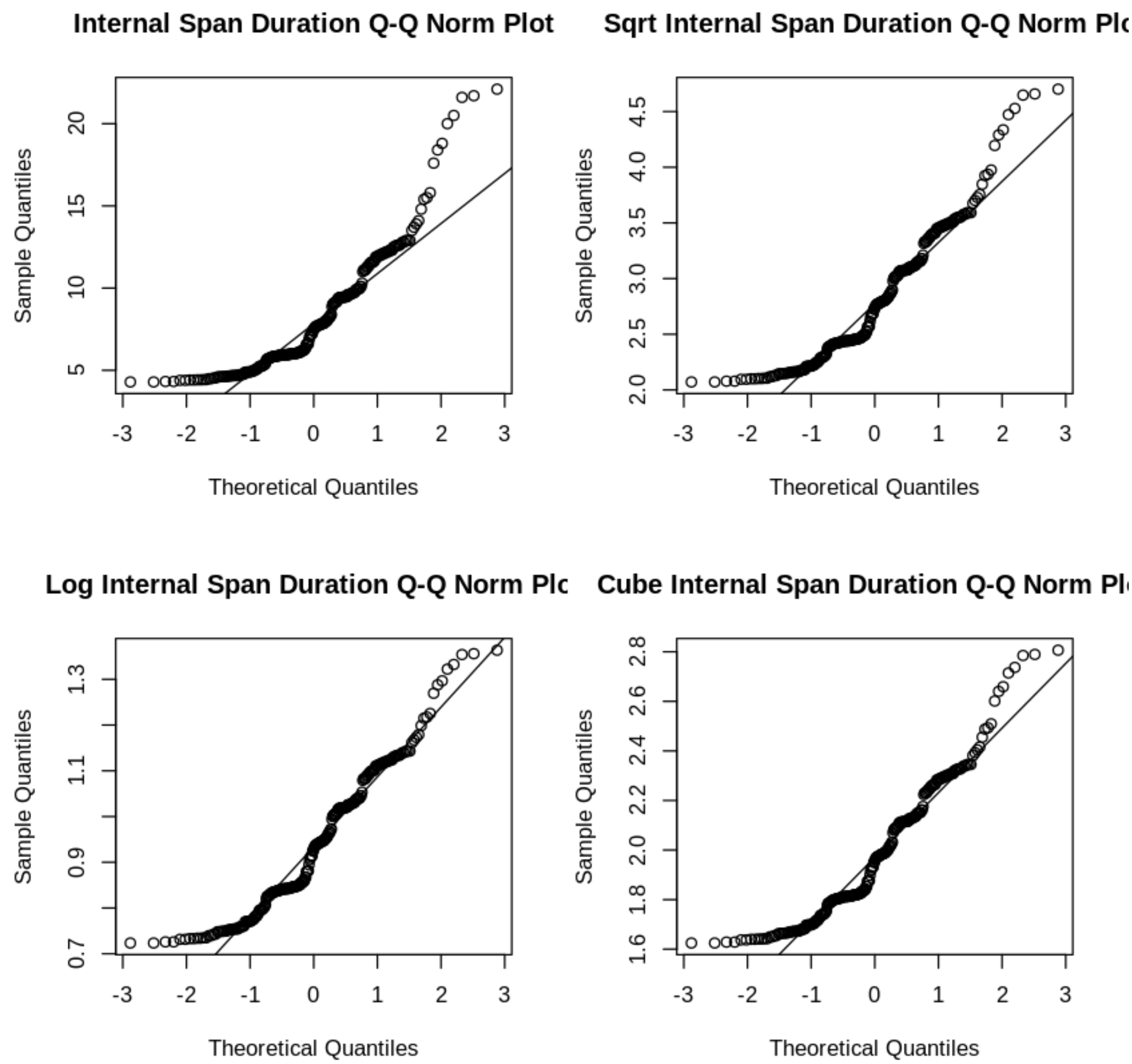
This result looks much better. However, we'll remove internal span outliers.

5.2.1 Data Transformations

5.2.1.1 Sqrt-Log-Cube Transformations



5.2.1.2 Q-Q Norm Sqrt-Log-Cube




```
    Shapiro-Wilk normality test

data:  log_iSpan$Duration
W = 0.94448, p-value = 3.486e-08
```

5.2.2 Box-Cox Transformation

Box and Cox (1964) developed a family of transformations designed to reduce nonnormality of the errors in a linear model. Applying this transform often reduces non-linearity as well, and heteroscedascity.

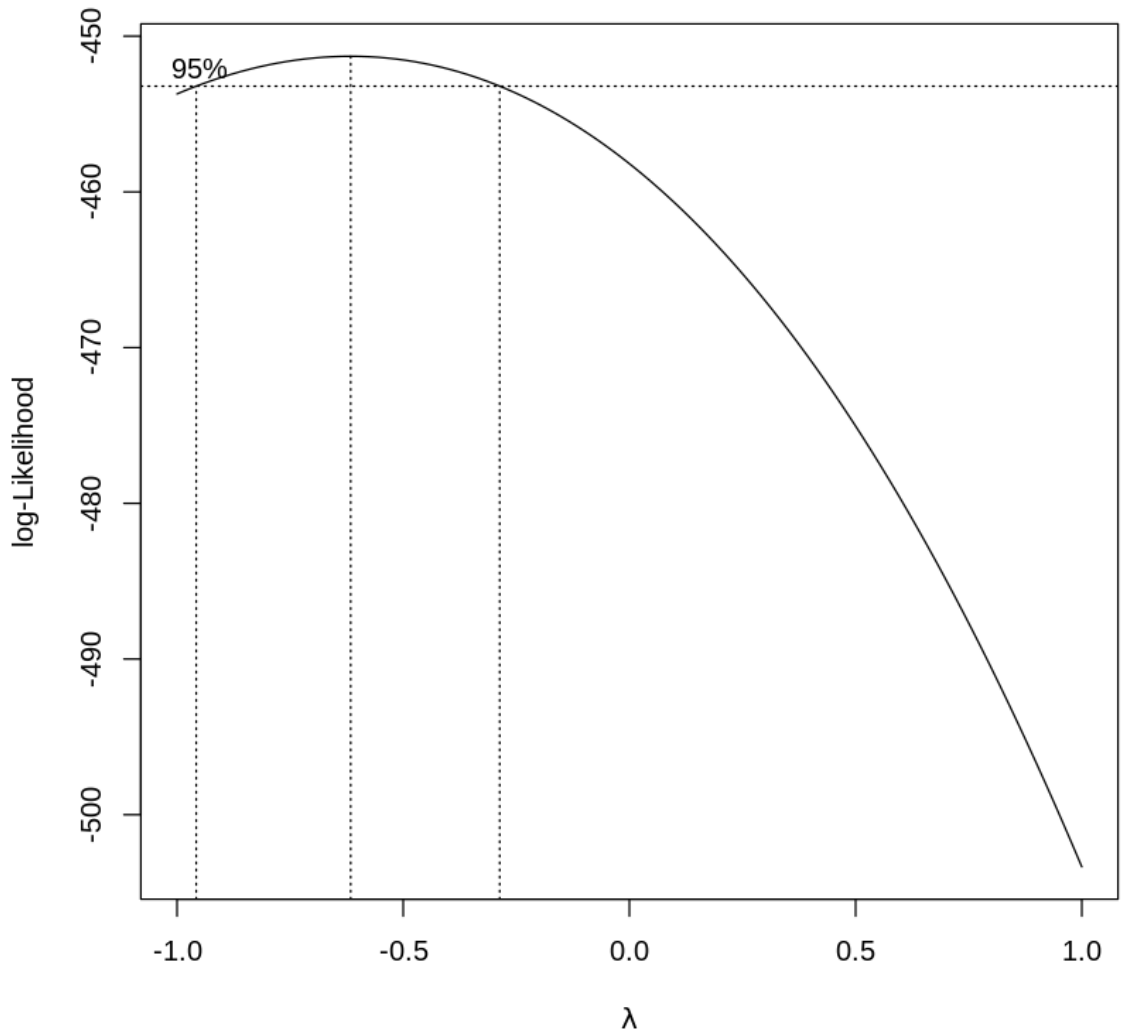
The idea is to transform the response variable Y to a replacement response variable $Y_i^{(\lambda)}$, leaving the right-hand side of the regression model unchanged, so that the regression residuals become normally-distributed. Note that the regression coefficients will also change, because the response variable has changed; therefore, the regression coefficients must be interpreted with respect to the transformed variable. Also, any predictions made with the model have to be back-transformed, to be interpreted in the original units.

The standard (simple) Box-Cox transform is:

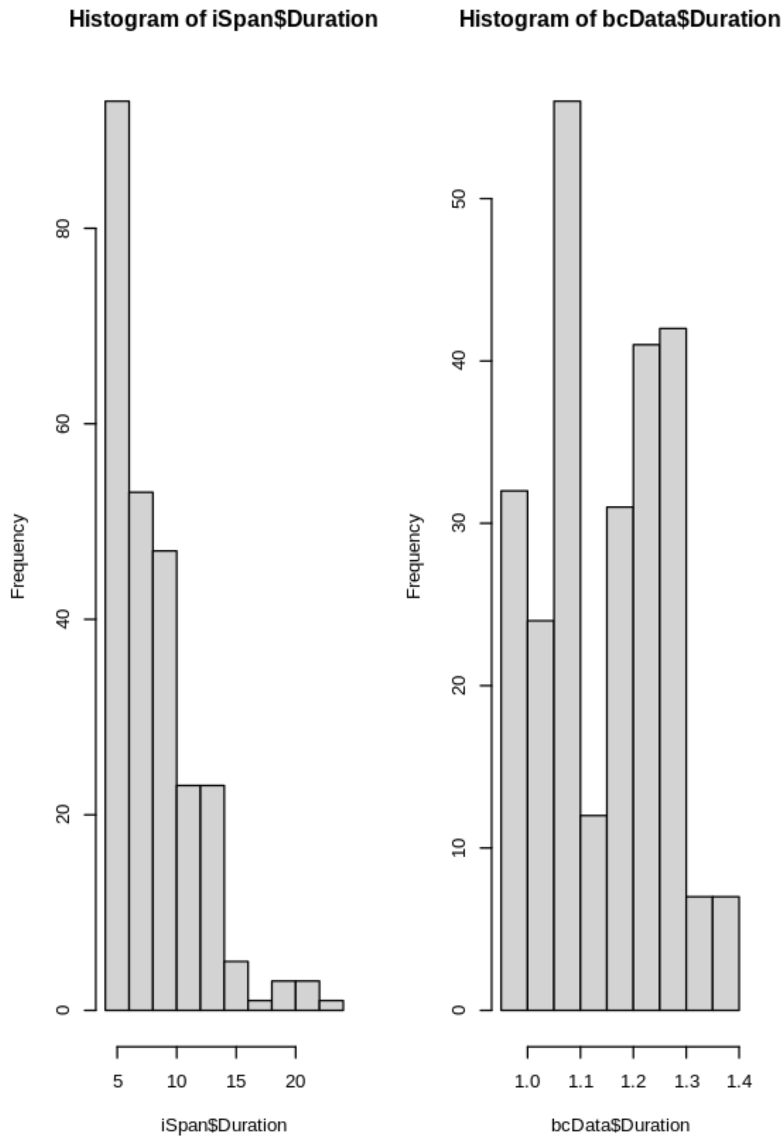
$$Y_i^{(\lambda)} = \begin{cases} \frac{Y_i^{\lambda}-1}{\lambda}, & (\lambda \neq 0) \\ log(Y_i), & (\lambda = 0) \end{cases}$$

Box, G. E. P., & Cox, D. R. (1964). An Analysis of Transformations. Journal of the Royal Statistical Society, Series B (Metholodogical), 26(2), 211-252.

http://www.css.cornell.edu/faculty/dgr2/_static/files/R_html/Transformations.html



95%
log-Likelihood
-450
-460
-470
-480
-490
-500
-1.0
-0.5
0.0
0.5
1.0
λ



5.3 Normality Testing of the Trasformation

5.3.1 Shapiro-Wilk

The null-hypothesis of this test is that the population is normally distributed. Thus, if the p value is less than the chosen alpha level, then the null hypothesis is rejected and there is evidence that the data tested are not normally distributed. On the other hand, if the p value is greater than the chosen alpha level, then the null hypothesis (that the data came from a normally distributed population) can not be rejected (e.g., for an alpha level of .05, a data set with a p value of less than .05 rejects the null hypothesis that the data are from a normally distributed population).

https://en.wikipedia.org/wiki/Shapiro–Wilk_test

```
    Shapiro-Wilk normality test

data:  bcData$Duration
W = 0.95906, p-value = 1.416e-06
```

With p-value of 2.852e-08 < 0.05 we reject the null hypothesis that the data are from a normally distributed population. But we'll also do a Q-Q Norm plot to visually see the results.

"if the p value is greater than the chosen alpha level, then the null hypothesis (that the data came from a normally distributed population) can not be rejected"

5.3.1.1 Shapiro-Wilk Testing Sqrt-Log-Cube

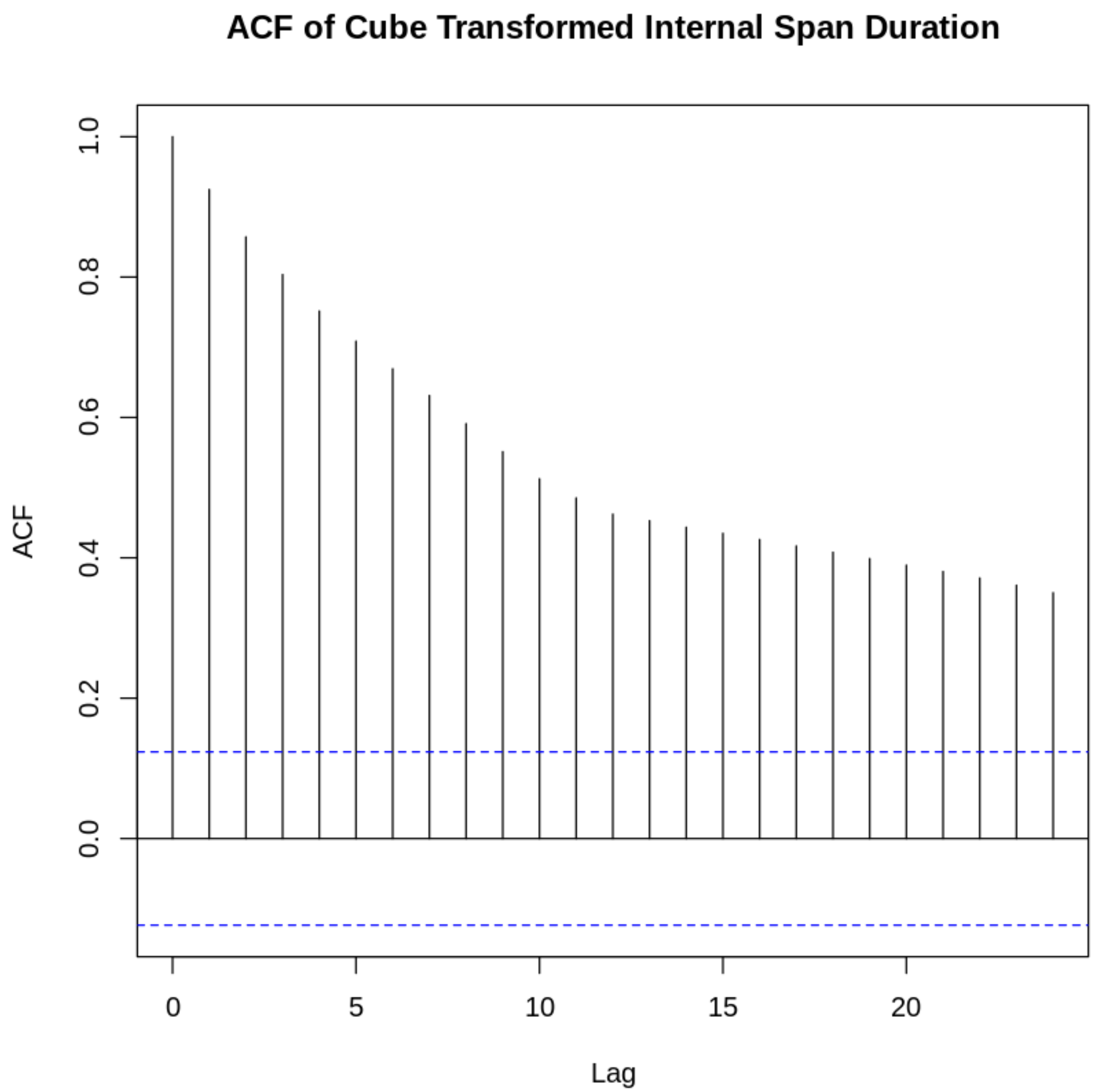
```
    Shapiro-Wilk normality test

data:  sqrt_iSpan$Duration
W = 0.92011, p-value = 2.208e-10

    Shapiro-Wilk normality test

data:  log_iSpan$Duration
W = 0.94448, p-value = 3.486e-08

    Shapiro-Wilk normality test

data:  cube_iSpan$Duration
W = 0.93233, p-value = 2.422e-09
```

The **cube transformation** seems to provide the best q-q plot fit. With a p-value of 0.3593 > 0.05 we fail to reject the null hypothesis and assume we now have a normal distribution.

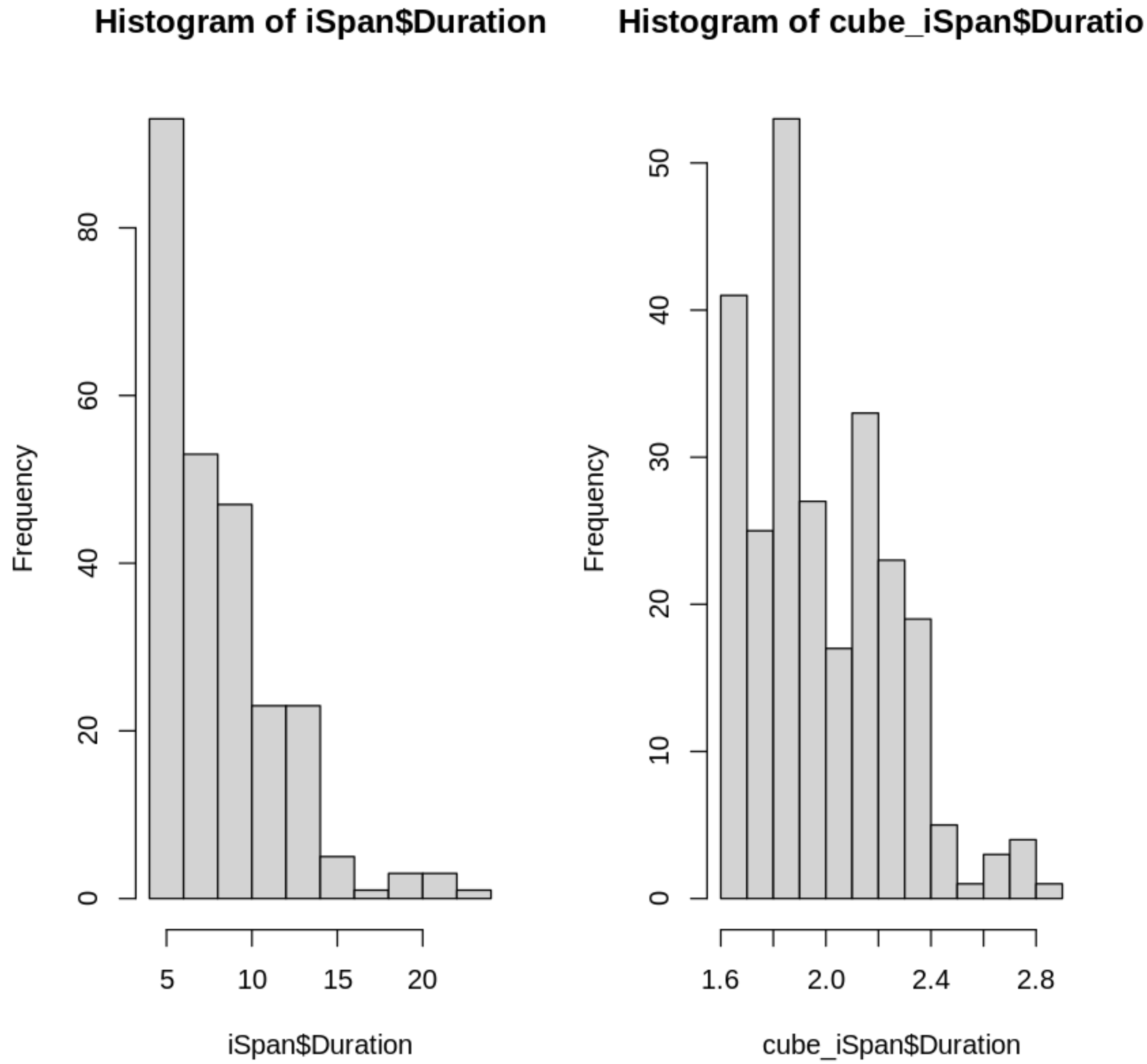
"if the p value is greater than the chosen alpha level, then the null hypothesis (that the data came from a normally distributed population) can not be rejected"

5.3.2 Autocorrelation

Autocorrelation plots are a commonly-used tool for checking randomness in a data set. This randomness is ascertained by computing autocorrelations for data values at varying time lags. If random, such autocorrelations should be near zero for any and all time-lag separations. If non-random, then one or more of the autocorrelations will be significantly non-zero.



The ACF indicates that the data is random since the results are near zero.



5.3.3 Hypothesis Testing

We will use a Student's t-Test to test the hypothesis on **normal** internal span data. Our mean is 500 ms (e.g. $\mu = 0.5$ seconds) and our null hypthesis is less than 500 ms.

```
    One Sample t-test

data:  x
```

```
t = 71.201, df = 251, p-value < 2.2e-16
alternative hypothesis: true mean is greater than 0.7937005
95 percent confidence interval:
 1.955862      Inf
sample estimates:
mean of x
 1.983448


    One Sample t-test

data:  x
t = 34.699, df = 251, p-value < 2.2e-16
alternative hypothesis: true mean is greater than 0.5
95 percent confidence interval:
 7.865648      Inf
sample estimates:
mean of x
 8.233611
```

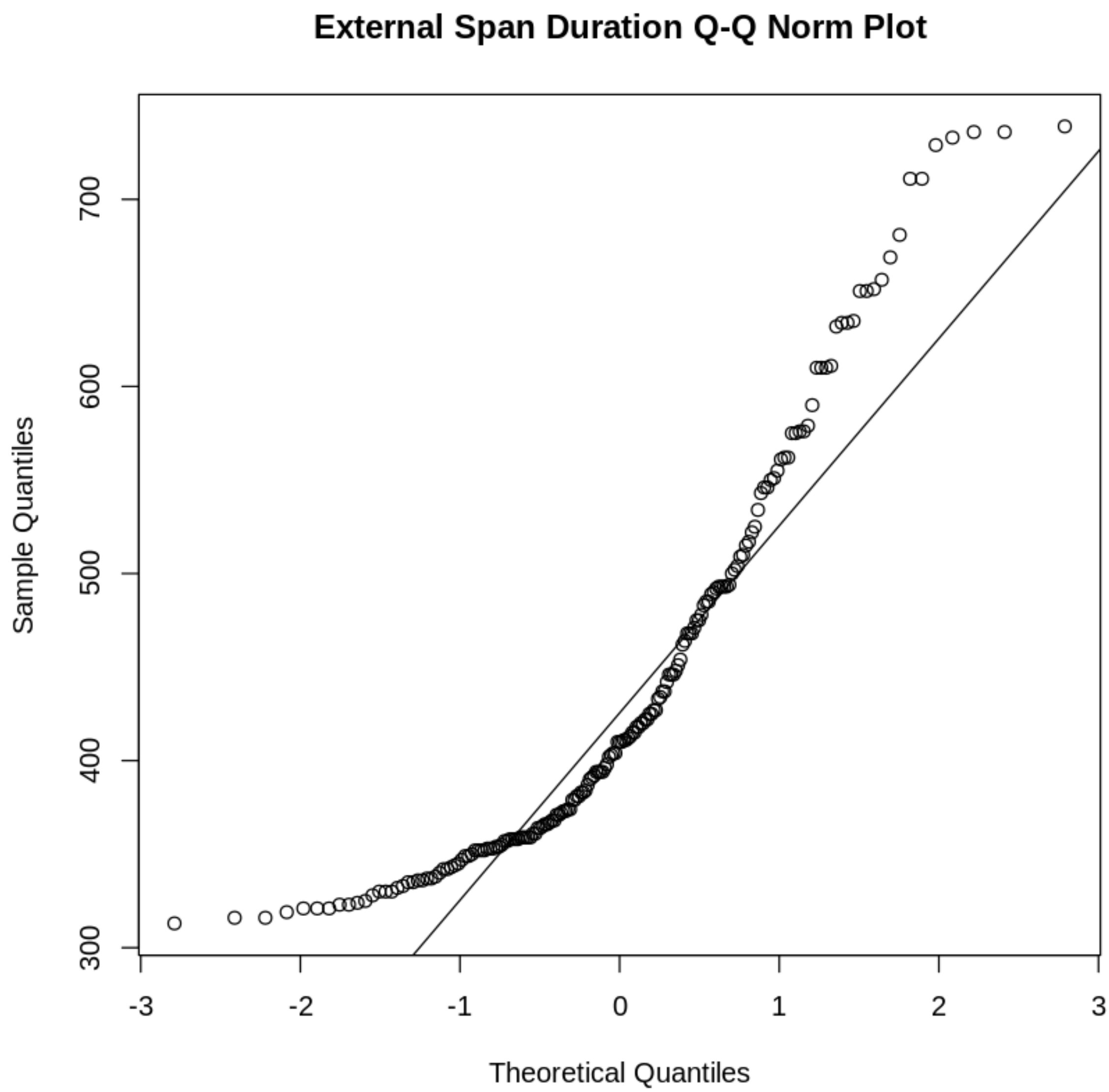
With a original and transformation with a p-value of $1 > 0.05$ we fail to reject the null hypothesis, i.e. we assume that latency will be less than 500 ms.

"If the p value is greater than the chosen alpha level, then the null hypothesis (that latency is $<$ 500 ms) can not be rejected"

5.4 External Data

5.4.1 Q-Q Norm Plot of "Clean" External Span Data

We'll look a the Q-Q Norm Plot and Shapiro-Wilk Test



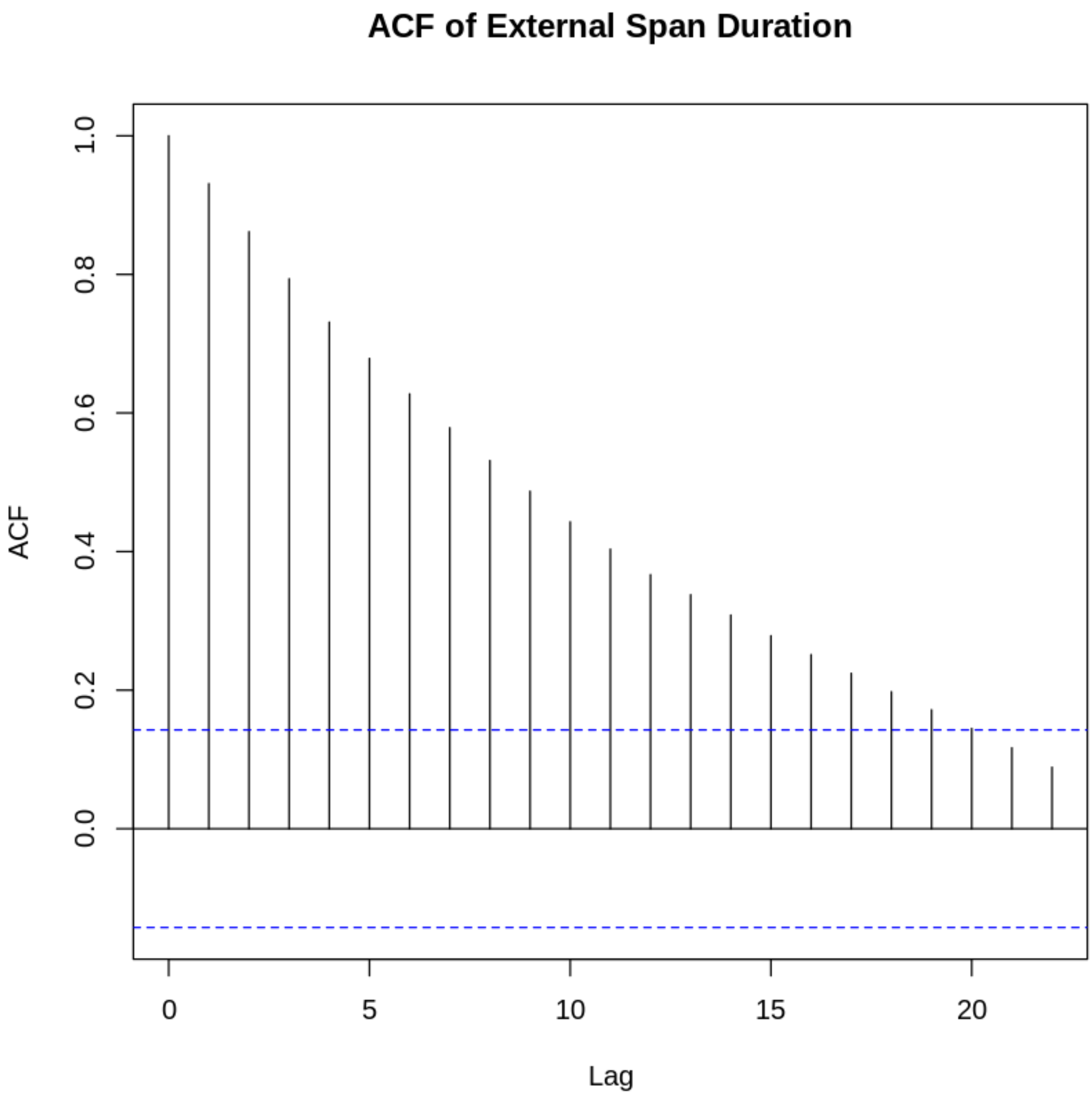
5.4.2 Shapiro-Wilk Normality Test

```
    Shapiro-Wilk normality test

data:  eSpan$Duration
W = 0.88201, p-value = 4.995e-11
```

With a p-value of $0.2878 > 0.05$ we fail to reject the null hypothesis, i.e. we assume that we have a normal distribution.

“if the p value is greater than the chosen alpha level, then the null hypothesis (that the data came from a normally distributed population) can not be rejected”

5.4.3 Autocorrelation



The ACF indicates that the data is random since the results are near zero.

5.4.4 Hypothesis Testing

We will use a Student's t-Test to test the hypothesis on external span data. Our mean is 500 ms (e.g. $\mu = 0.5$ seconds) and our null hypthesis is less than 500 ms.

```
    One Sample t-test

data:  x
t = 56.515, df = 188, p-value < 2.2e-16
alternative hypothesis: true mean is greater than 0.5
95 percent confidence interval:
 427.8283      Inf
sample estimates:
mean of x
 440.7037
```

With a p-value of $0.1336 > 0.05$ we fail to reject the null hypothesis, i.e. we assume that 500 ms can be maintained for external service requests.

"If the p value is greater than the chosen alpha level, then the null hypothesis (that latency is $<$ 500 ms) can not be rejected"

6 Observations

6.1 General Discussion of Normality

It was required to separate external data from internal to establish normality of the data samples. The internal data set required transformation to establish normality, while the external data did not require a transformation.

6.2 Hypothesis Results

Hypothesis testing using the Student's t-Test indicates that latency constraints of 500 ms can be maintained internally and external. However, serveral external samples were greater than 500 ms. This is most likely due to the non-deterministic nature of internet (e.g. http) requests. Within the internal environment, data is directly routed between microservices within the Docker environment within a private network. The data shows that a container based microservice architecture can meet the requirement; however, care must be taken to manage processing per container that may increase container response times.