



Doing Data Science Right in Excel-Pervasive Utilities

Eina Ooka

September 25, 2018



The Energy Authority (TEA)

- **The Energy Authority** serves public utilities nationwide for trading and analytics.
- Analytics team provides forecasting and analysis including:
 - Load and market forecasts
 - Portfolio risk and hedge analysis
 - Congestion and transmission analysis
 - IRP studies and many more...
- A few dozen analysts in the team



TEA Analytics ~ a few years ago ~

- Excel clutter
 - Accumulation of **errors** over time. Hard to detect them in enormous multi-tab models.
 - Duplicates of **a same model for different versions** (e.g., for different customers). When making changes, same changes needs to be made multiple times.
 - **No version controls**. Accumulation of badly named similar excel files and many no-longer used scripts and calculations.
 - For large dataset, it was slow.
- Multiple languages were used in addition to excel
 - VBA, R, matlab, python, ...
- No clear guideline for analytics quality control

My Goals

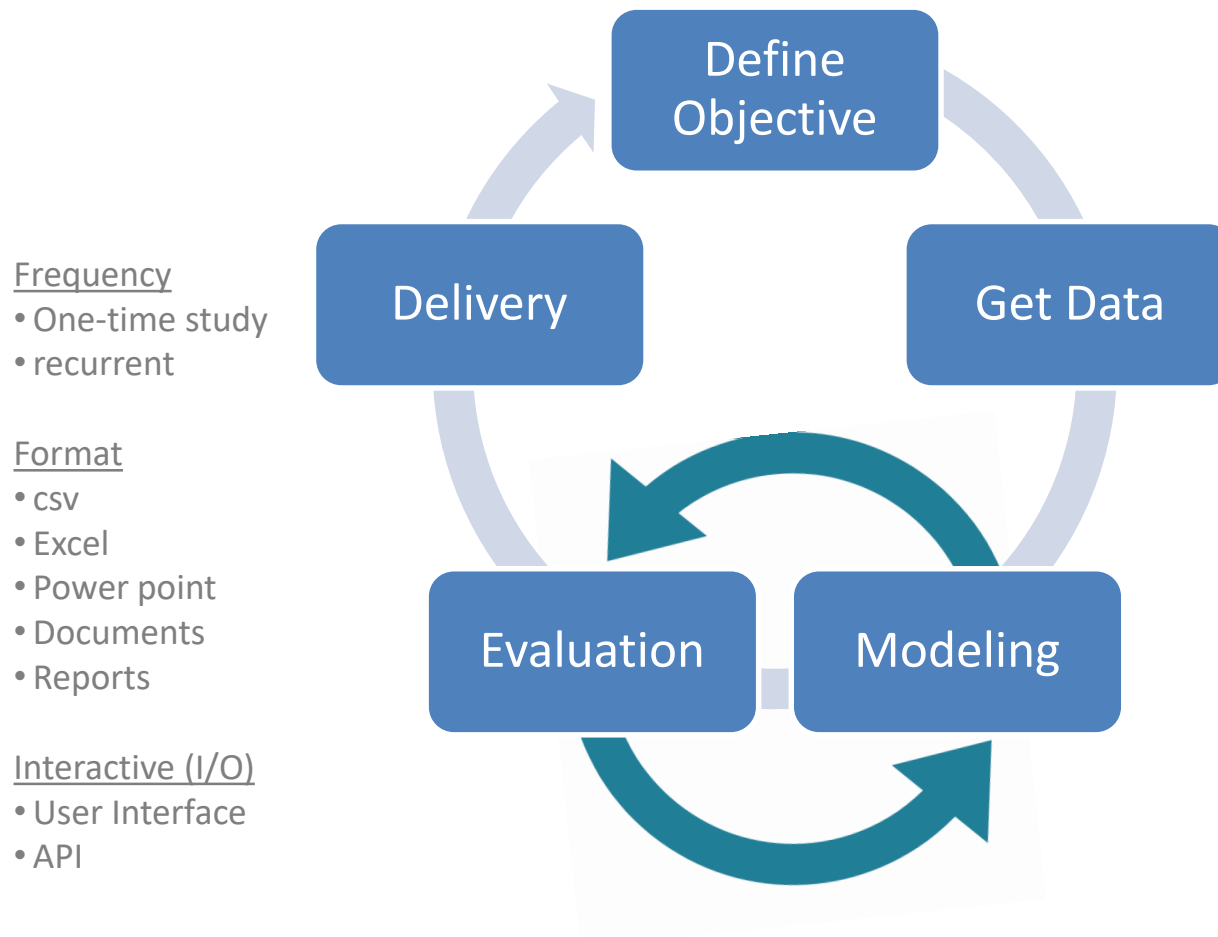
- Build out a data science platform that enables:
 - Analytics Quality Control
 - Efficient and rapid development of a solution into delivery
 - Encourage collaboration

Terminology

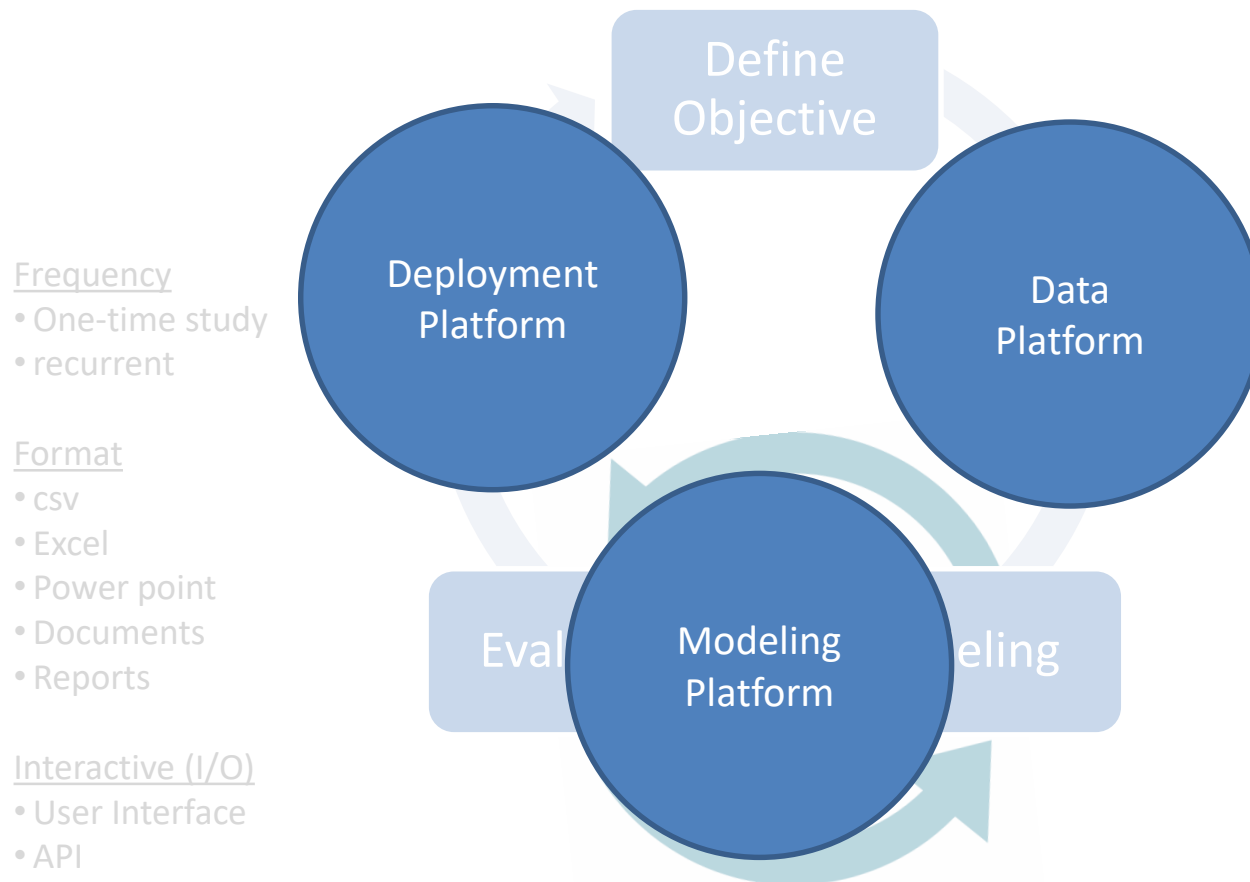
- A platform is the environment on which computer programs can run. For development process, it can be a constraint or an assistance.
- A computer program is written in a programming (or scripting) language. In our case, it is a model and an analysis written in scripting languages.

What's Needed in Data Science Process?

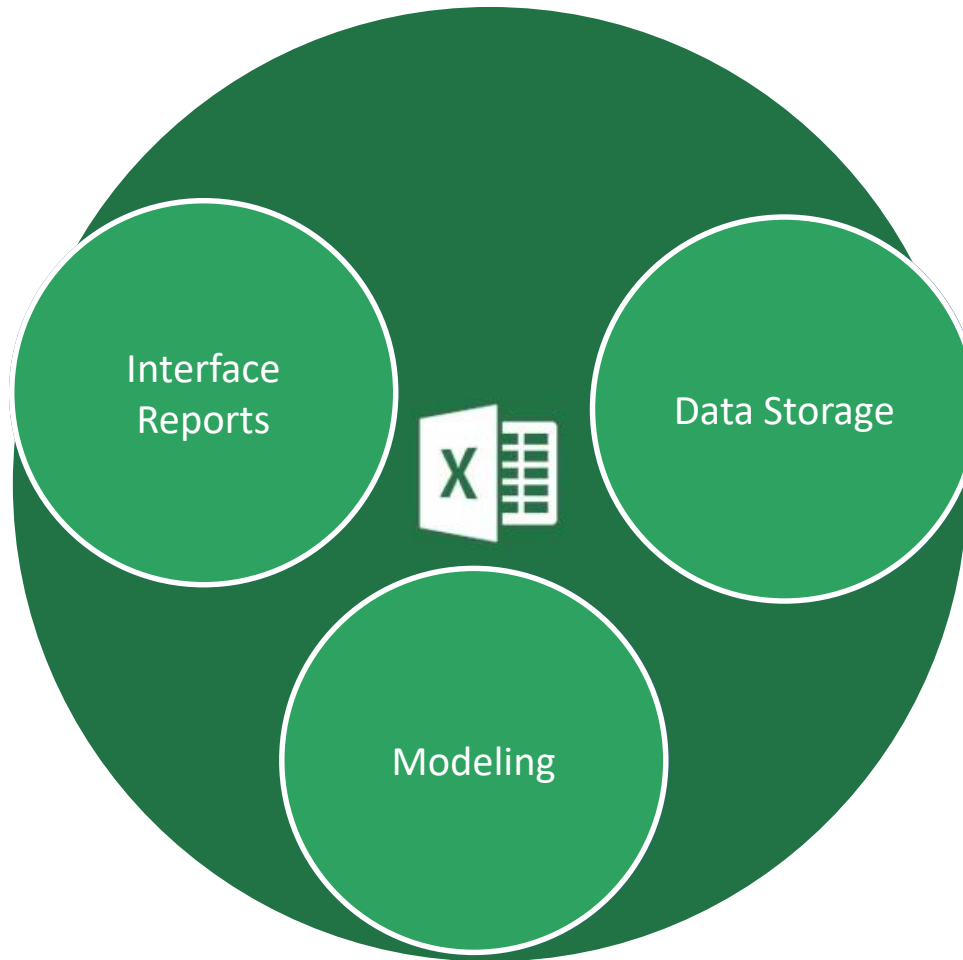
Data Science (Analytics) Lifecycle



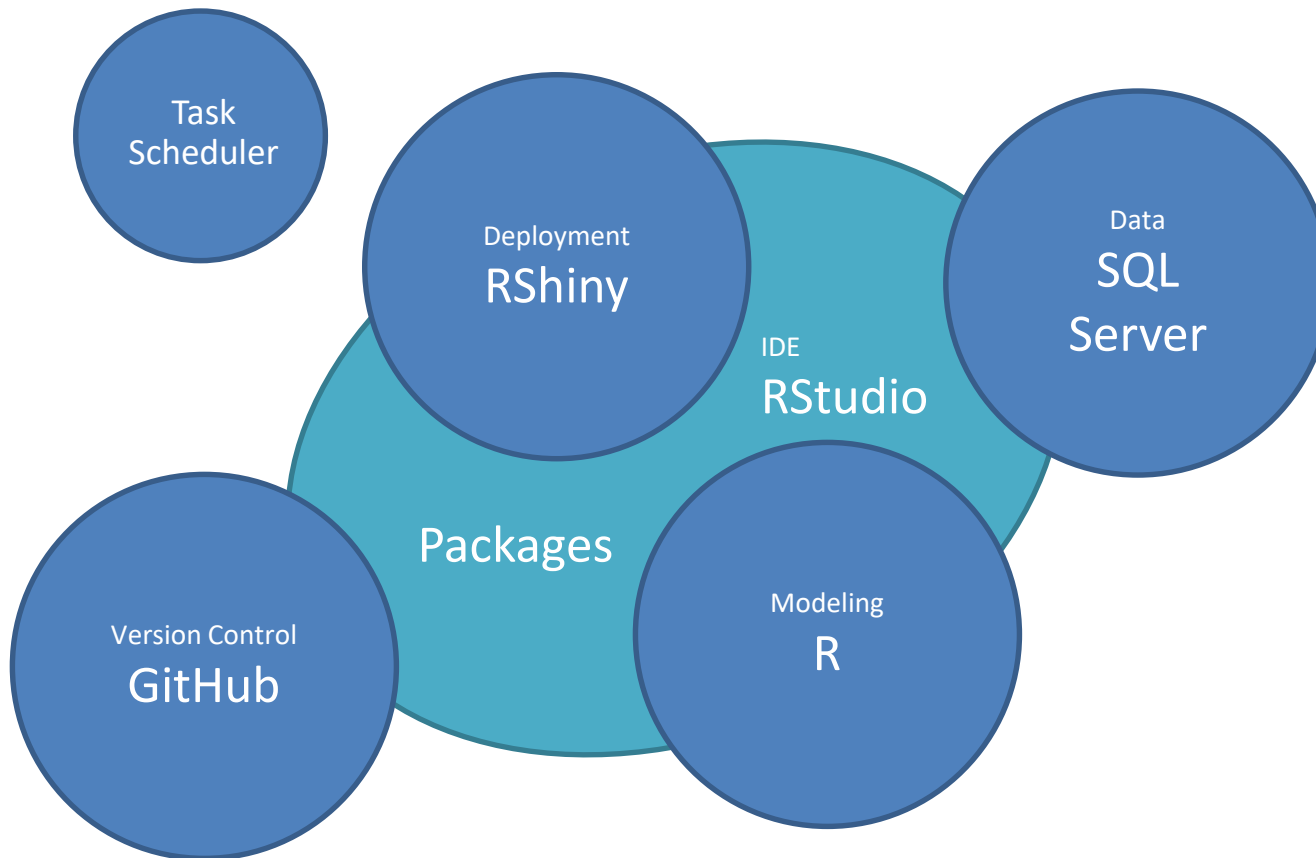
Data Science Platforms



Excel - Jack of All Trades



Data Science Platforms



Excel or non-Excel?

Use Excel if

- Summary data
- Model- and data-light
- Share with others



- A quick task where a jack-all-trades is all you need.

Do Not Use Excel if

- Needs scripts
- Model-heavy
- Data-heavy
- Recurring Task



- Need a team of masters

Analytics Quality Control

Aspirations for Model Design

1. Accurate
2. Robust
 - Do outliers affect the results too much?
3. Parsimonious
 - Is the model overly complicated?
4. Explainable
 - Simple understandable model is better than black box.
5. Consistency
 - Is the behavior consistent over time? Backcast.

Aspirations for Model Design

6. Adaptable

- Can we use parts of the model on other projects?

7. Scalable

- Can we scale up the project to includes many other nodes, customers, etc...?

8. Accessible

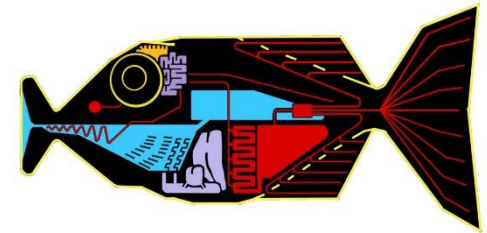
- Can other people understand and run the model?

9. Reproducible

- Can we produce the same results later?

Common Language

- Efficiency gain
 - Collaborations between analysts
 - Maintenance of a model over time
 - Re-use and share codes
- Protect against loss of work
 - Others can easily inherit other's models
 - Hedge against a sudden leave of an analyst

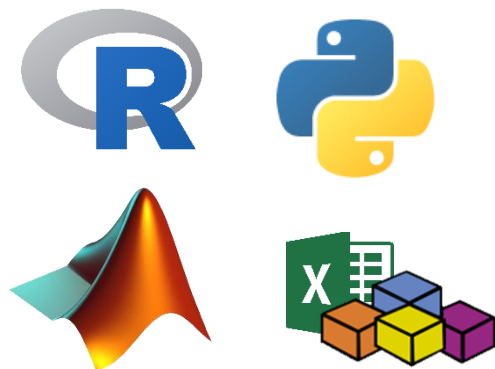


Evaluations

MORE QUALITATIVE THAN QUANTITATIVE

Modeling Platform

- We evaluated several scripting languages.
- Quantification by scoring was helpful but not the determining factor. For us, decision was more subjective.
- Our choice was R.



		R		Excel plus VBA
	importance	Score	User Friendly data analysis, statistics and graphical models	Score
		344		239
Analysis & Computations		61		26
Solid data analysis and statistics toolset	3	5		2
Advanced visualizations	3	5		3
interactivity	1	4	HTML widget	4
Advanced data science tools	2	4		1
Speed of computations	2	4	Not optimized for speed	2
Multithreading	1	3	Hard to use.	1
Non-numeric data handling	2	4		
Data Handling		56		19
Driver to basic type of data sources (SQL, excel)	3	5		4
Driver to non-standard data sources (NoSQL, Spark)	1	4	Limited Capability	1
Web scraping capability	2	4		1
Working with large dataset (~a few gigabytes)	2	4	Need discipline in good coding	1
Efficiency in memory usage	2	3	Need discipline in good coding	1
Data cleaning and imputations	3	5		
Developments		93		76
Readability	2	3		2
Workable IDEs	3	5	Rstudio	5
Version Control	3	5	Super easy with	1

Data Platform

Dedicated to Analytics

- We evaluated on-premise and on-cloud databases including types of SQL & NoSQL.
- How does it perform w.r.t. R?
 - DBI compliant? Flexibility of NoSQL?
- Our choice was SQL Server.



INTERFACE: R		ON-PREMISE						OFF-PREMISE		
Technology		Flatfile (rds)	Flatfile (csv)	SQL Server 2008	SQL Server 2016	Oracle (with RODBC)	MongoDB 3.4	Azure SQL Database	Azure DocumentDB (mongoDB API)	mongoDB 3.0
Server Location		Jax	Jax	Jax	Jax	Jax	Jax	Cloud (East)	Cloud (East)	Cloud (West)
Client Location		Jax	Jax	Jax	Jax	Jax	Jax	Jax	Jax	Jax
Data	Function									
mtcars df (32 x 11)	create a table	0.22	0.14	0.80	0.11	4.09	0.14	2.05	1.43	0.47
	pull all	0.19	0.22	0.01	0.02	0.03	0.02	0.25	0.25	0.17
	filter 1	NA	NA	0.02	0.01	0.04	0.02	0.22	0.23	0.15
	filter 2	NA	NA	0.03	0.01	0.01	0.02	0.22	0.22	0.15
	append data	NA	NA	0.03	0.10	3.99	0.03	0.72	0.28	0.10
	update entries	NA	NA	0.02	0.00	0.00	0.02	0.24	0.21	0.08
	remove rows	NA	NA	0.01	0.00	0.01	0	0.21	0.22	0.06
	aggregation	NA	NA	0.00	0.01	0.00	0.01	0.21	NOT AVAILABLE	0.07
	delete table	NA	NA	0.01	0.01	0.24	0.02	0.23	0.50	0.14
	Data size (MB)	0.002	0.002	0.008	0.008			0.008	0.000	
	Index size (MB)	NA	NA	0.008	0.008			0.008	0.008	
nycflights df (336,776 x 19)	create a table	6.8	28.0	954	664	435	30	11360	2523	83
	bulk insert	NA	NA	29	30		NA	NA	NA	NA
	pull all	1.5	15.0	9.0	9.7	16.7	44.6	23.2	331.9	10.0
	filter	NA	NA	0.5	0.2	0.1	0.7	0.3	1.0	1.4
	sort	NA	NA	10.4	10.4	16.7	Failed.	17.6	301.9	Failed.
	aggregation	NA	NA	0.8	0.2	0.1	1.2	0.3	NOT AVAILABLE	0.5
	Data size (MB)	6.8	35.8	38.5	37.4	35.0	31.7	43.2	307.9	174.7
	Index size (MB)	NA	NA	6.06	6.10		5.41	6.06	47.7	19.4
hadoopfs - original	create a table	NA	NA	NA	NA	NA	NA	NA	NA	NA

Delivery Format

- If we simply deliver flat files or reports, no need for deployment platform.
 - R-markdown has made R reporting extremely easy and visually appealing.
 - There are numerous r packages that enable saving results in Excel.



- Deployment platform is necessary when a deliverable is a model with inputs/outputs.



Frequency

- One-time study
- Recurrent

Format

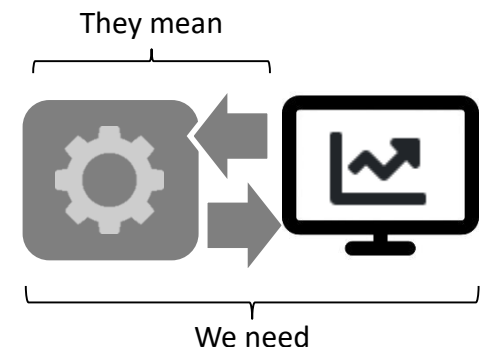
- csv
- Excel
- Power point
- Documents
- Reports

Interactive (I/O)

- User Interface
- API

Data Science Deployment

- When non-utility data scientists say “deployment,” they usually mean:
 1. The model is turned over to software developers to translate into a production language.
→ We don’t have software developers to spare...
 2. Deploy models on their own as APIs.
→ We don’t have web developers to spare...
- How do we create a user interface for our models without going back to Excel or relying on IT developers?



Deployment Platform

- We evaluated
 - Various “Data Science Platforms.”
 - There were a lot fewer numbers to evaluate in 2015.
 - MS Azure
 - They were starting to push Azure ML and other services.
 - Various BI tools
 - R Shiny Server



- We chose R Shiny Server.
 - Shiny (r package): HTML wrapper and generate apps on a desktop.
 - Shiny Server: web service for hosting shiny apps.

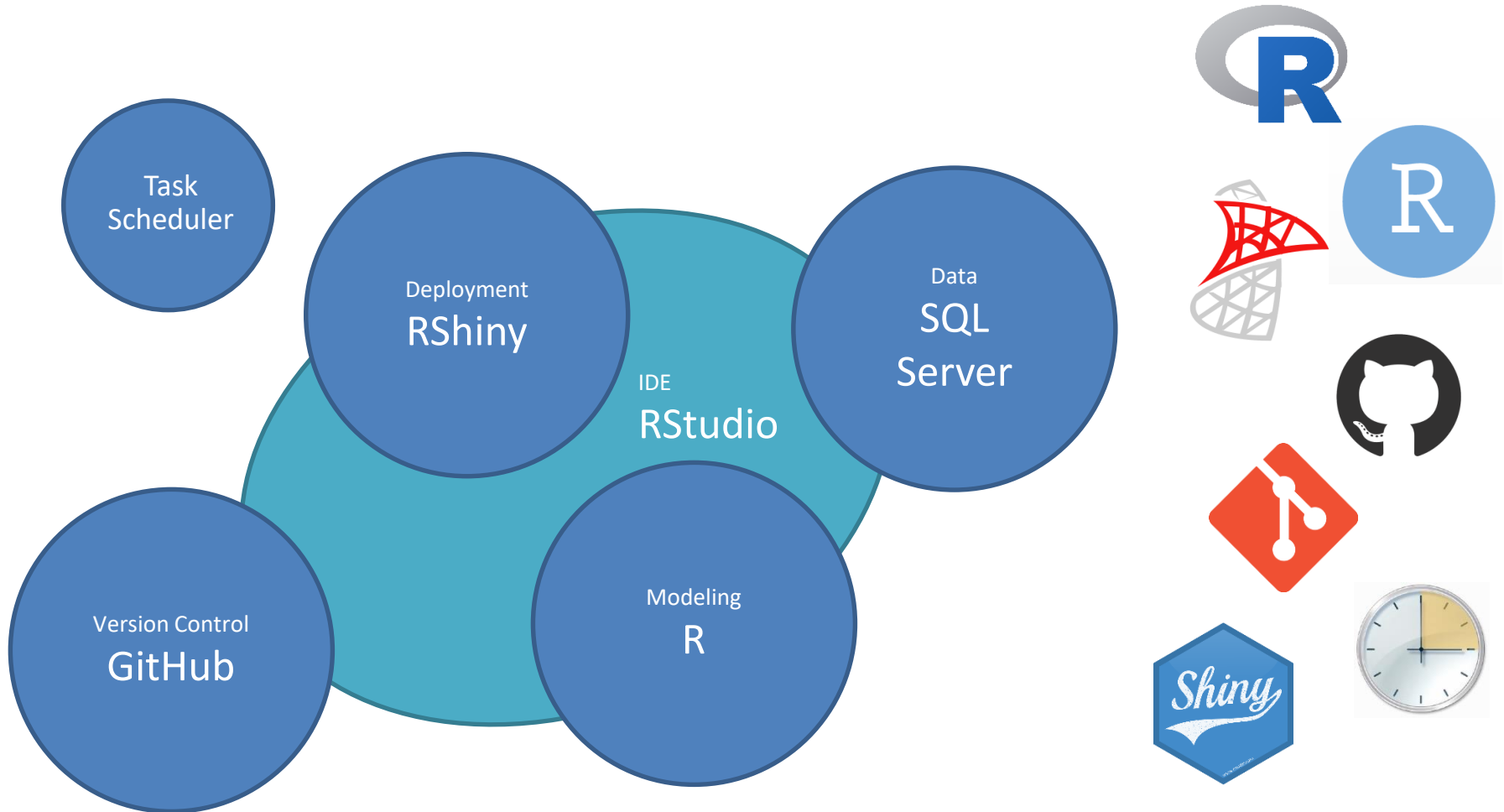
Vision

Deployment Platform ~~Evaluation Criteria~~

- **User Interface**
 - Web-based or installed app?
 - Visualization capability and its flexibility? Can we show r plots?
 - Can we modify inputs and re-run the model from the interface?
 - Download data or reports from dashboard.
 - Interactive plots
 - display on mobile devices
- **Ease of deployment**
 - How much IT help do we need at the time of deployment?
- **Data**
 - Can it connect to TEA's internal databases?
 - Is there a limit to the size of data it can handle?
- **Accessibility**
 - Authentication? Differentiation in authentication levels?
 - Is the dashboard accessible from internal and external users?
- How many people can access the application concurrently?
- **Scalability**
 - How scalable? If so, does the price reflect it?
- **Languages**
 - R
 - Python
- **API & Web Services**
 - Can we call models from BI?
- **Other functions**
 - Scheduled runs?
 - Version Control?
 - Collaboration?
- **Technical Support**
 - Is technical support available?
 - Are there online resources?
 - How responsive are they?



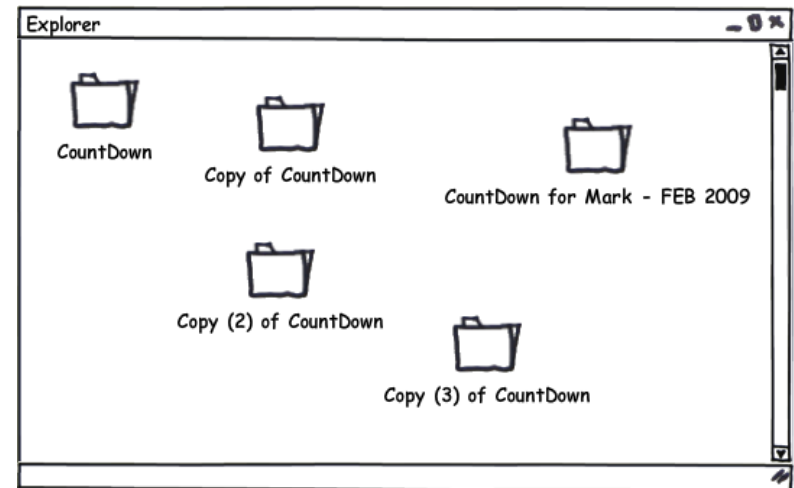
Data Science Platforms





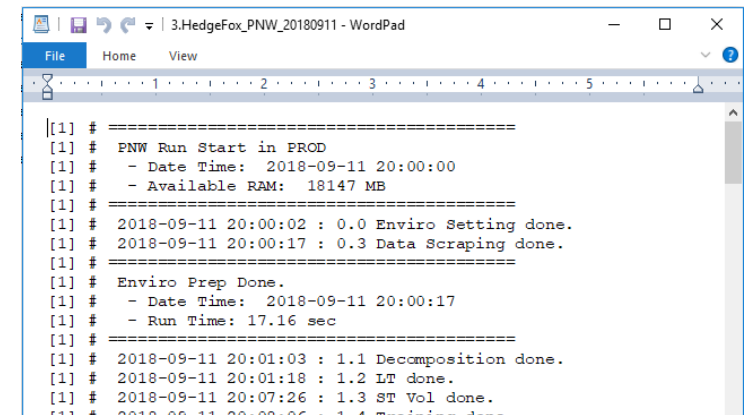
Source Control

- Git for
 - Source Control
- GitHub for
 - Managing Repositories
 - Project Management
 - Collaboration
 - Issue tracker
- Use of Git, together with Shiny was a game changer for us.
 - Eliminated accumulation of similarly named files or commented-out old scripts.
 - Easy differentiation between multiple (Dev, PROD, ...) environments.



R Other Supporting Tools

- RStudio IDE (Integrated Development Environment)
 - In recent years, it has integrated many features necessary for efficient data science work.
- Task Scheduler with Batch Tasks
 - You can add arguments
 - Don't forget to output a run log.
- Setting up Dev, Staging and PROD environments.
- Set naming conventions and coding practice guidelines.
- Data Camp for new users within the group.



```
[1] # =====  
[1] # PNW Run Start in PROD  
[1] # - Date Time: 2018-09-11 20:00:00  
[1] # - Available RAM: 18147 MB  
[1] # =====  
[1] # 2018-09-11 20:00:02 : 0.0 Enviro Setting done.  
[1] # 2018-09-11 20:00:17 : 0.3 Data Scraping done.  
[1] # =====  
[1] # Enviro Prep Done.  
[1] # - Date Time: 2018-09-11 20:00:17  
[1] # - Run Time: 17.16 sec  
[1] # =====  
[1] # 2018-09-11 20:01:03 : 1.1 Decomposition done.  
[1] # 2018-09-11 20:01:18 : 1.2 LT done.  
[1] # 2018-09-11 20:07:26 : 1.3 ST Vol done.  
[1] # 2018-09-11 20:08:06 : 1.4 Machine done
```


Case Study 1

AURORA MODELING

- PREVIOUS CUT&PASTE METHOD
- CURRENT ONE-CLICK REPORTING METHOD

Before...

AURORA Reporting Task

- StudyNote
- StudyNote - Base Case - 1st Run
- StudyNote - Base Case - 2nd Run
- StudyNote - Base Case - 3rd Run
- StudyNote - Base Case - 3.5rd Run
- StudyNote - Base Case - 4th Run
- StudyNote - Base Case - 5th Run
- StudyNote - Base Case - 5.5th Run
- StudyNote - Base Case - 6th Run

...

Microsoft PowerPoint Pre...

Microsoft PowerPoint Pre...

Microsoft PowerPoint Pre...

Microsoft PowerPoint Pre...

Microsoft PowerPoint Pre...

Microsoft PowerPoint Pre...

Microsoft PowerPoint Pre...

Microsoft PowerPoint Pre...

Assumptions

5.5, 5.6, 5.7

- PIRA Henry NG price
- Renewable additions, at least to meet RPS.
 - Take out max addition limit for Wind and Solar.
- CA mid carbon price: \$50 by 2030
- Table setup corrected.
- Clean Power Plan, but no CA Carbon Cap
- CPP Limit interpolation from 2022 (at 2012 level) and 2030 (at goal).

Solar overnight cost reduced to

5.5 CA Resource Additions and Retirements

CPP & 50% CA RPS & CA Mid-Carbon Price

mission

2012

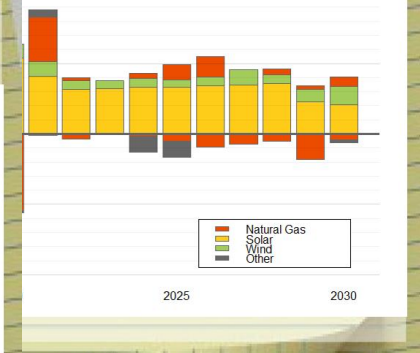
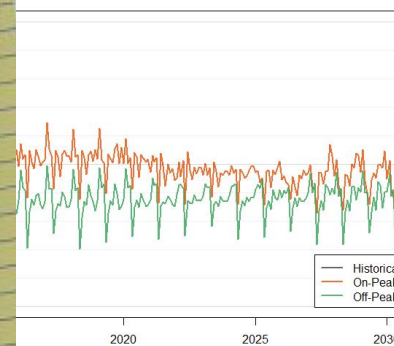
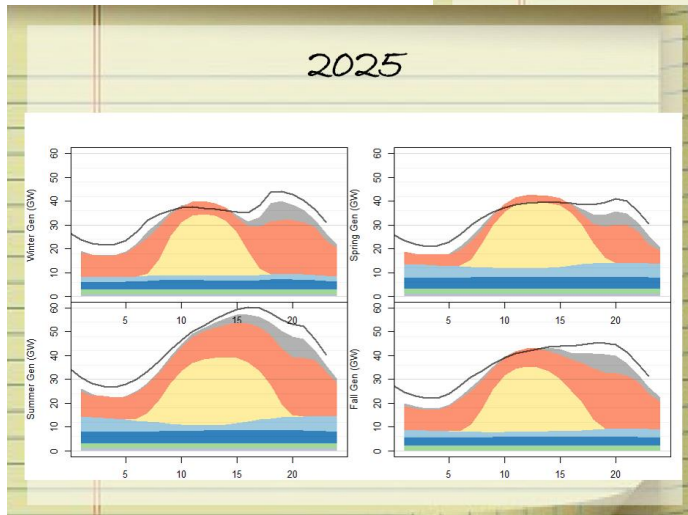
2

), and

same rate

Mid-C Implied Heat Rate

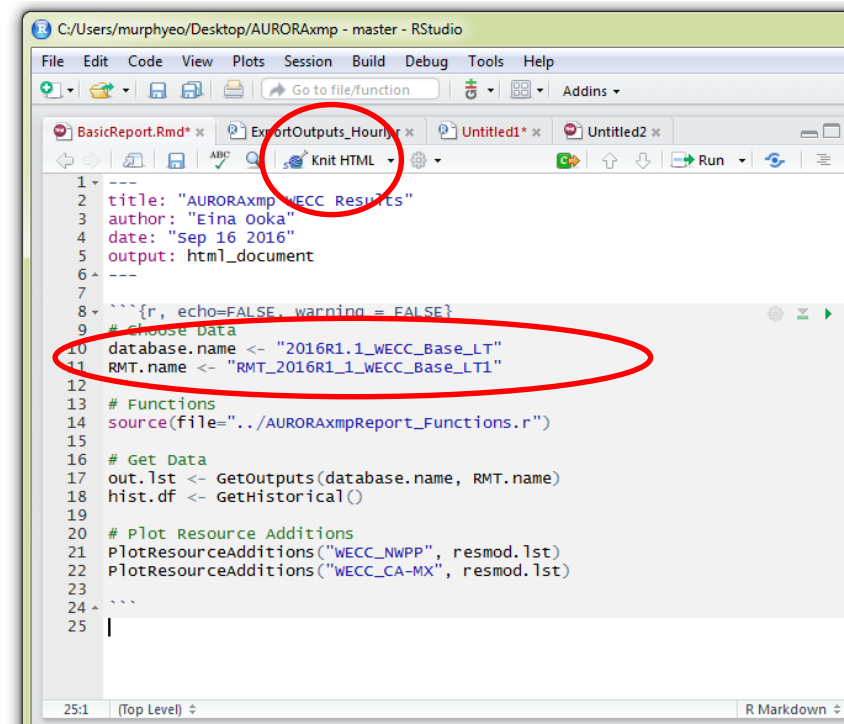
h Clean Power Plan



Current Reporting Process

AURORA Reporting Task

- Run the cost-production model and export all outputs into a database.
- Open R. Update the database names and “knit.”
- Done.



```

1 ---
2 title: "AURORAXmp WECC Results"
3 author: "Eina Ooka"
4 date: "Sep 16 2016"
5 output: html_document
6 ---
7
8 ```{r, echo=FALSE, warning = FALSE}
9 # Choose Data
10 database.name <- "2016R1.1_WECC_Base_LT"
11 RMT.name <- "RMT_2016R1.1_WECC_Base_LT1"
12
13 # Functions
14 source(file="../AURORAXmpReport_Functions.r")
15
16 # Get Data
17 out.lst <- GetOutputs(database.name, RMT.name)
18 hist.df <- GetHistorical()
19
20 # Plot Resource Additions
21 PlotResourceAdditions("WECC_NWPP", resmod.lst)
22 PlotResourceAdditions("WECC_CA-MX", resmod.lst)
23
24 ^
25 |
  
```

Model Specifications

Load Growth
Resource Additions
RPS
Resource Type Outputs
Resource Stack
Resource Type Cash Flow
Market Prices
CPP Constraint
Transmission

AURORAxmp WECC Results

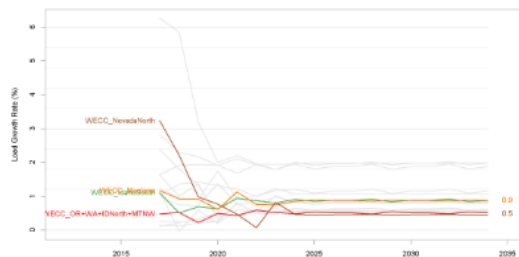
Eina Ooka
May 3 2016

Model Specifications

- Database: 2016R1.1 WECC_Base_LT
- Run ID: 2016R1 WECC_Base_LT
- Run Date: 5/2/2016 3:13:42 AM
- Run Time: 1491.1589 min
- AURORAxmp: x64 version of AURORAxmp 12.1.1015
- File Name: Project: D:\016_AURORAxmp_1_LT.apz
- Sampling: 1st and 3rd Week: Mon, Wed, Fri, Sun; Every 3rd Hour
- Inflation rate: 1.7%

Load Growth

NWPP CA RMRG SRSG Canada



Resource Additions

Cumulative Resource Additions

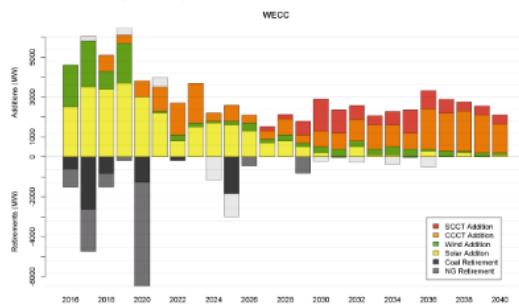
Name	WECC_CA-MX	WECC_NWPP	WECC_RMGR	WECC_SRSG
CCCT	1410	15200	0	8400
SCCT	7590	0	0	2300
Solar	11100	8000	1400	8200
Storage	1125	0	0	0
Wind	900	9200	1505	600

Cumulative Resource Retirements

Fuel	WECC_CA-MX	WECC_NWPP	WECC_RMGR	WECC_SRSG
Coal	115	6631	1127	1375
FO	0	5	10	0
NG	10670	2069	361	277
PS	1325	0	0	0
UR	2300	0	0	0
WAT	1	47	0	0

WECC NWPP CA-MX RMRG SRSG

Boardman, Centralia and a part of Colstrip is scheduled to retire between 2020 and 2026.



RPS

Because of imports and exports between states, renewable generation output percentages within each state may not coincide

Model Specifications

Monthly Prices
Hourly Hub Prices
Hourly Resource Types

AURORAxmp WECC Hourly Results

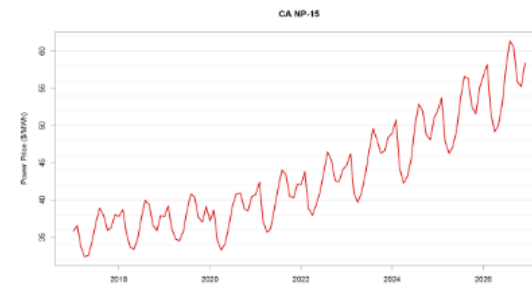
Eina Ooka
May 4 2016

Model Specifications

- Database: 2016R1.1 WECC_Base_Hourly2017to2026
- Run ID: 2016R1 WECC_Base_Hourly
- Run Date: 5/3/2016 8:09:22 PM
- Run Time: 149.5805 min
- AURORAxmp: x64 version of AURORAxmp 12.1.1015
- File Name: Project: D:\016_AURORAxmp_1_Hourly.apz
- Inflation rate: 1.7%

Monthly Prices

NP15 Mid-C

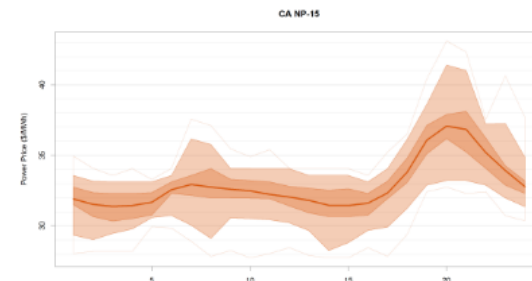


Hourly Hub Prices

NP15 Mid-C

2017 2021 2026

Spring Summer Fall Winter

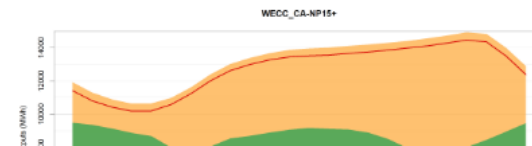


Hourly Resource Types

NP15 OH+WA+ID+North+MTNW

2017 2021 2026

Spring Summer Fall Winter



Case Study 2

PORTFOLIO RISK MODEL

- PREVIOUS EXCEL & VBA FRUSTRATION
- CURRENT ALL-R DISCIPLINE

Portfolio Risk Model

Before...

- There were Stochastic Portfolio Model written in VBA with Excel Interface.
- Produced large excel and csv files.
- We had copied of similar models with some customizations, for each customer.
- We saved copies of old models in Archive folder.

The screenshot shows an Excel spreadsheet titled 'Book1 - Excel' with the following tables:

Run Parameters	
Number of Iterations	1000
Simulation Start Date	2/1/2017
Simulation End Date	12/31/2020
Current Water Year Lower Limit	0%
Current Water Year Upper Limit	100%
Fiscal Year Begin Month	Jan
Annual Load Growth Scaling	0%
Forward Gas Price Scaling	0%
Narrowed Slice Distribution	TRUE

Customer / File Parameters	
Customer	
Parent Risk Model Directory	O:\WC Risk\Stochastics\
Parameter File Name	_out_20160318.xml
Actuals File Name	_actuals_20160318.csv
Stochastic Data File Name	RawOut_20170106.csv
Source Code File	Module1.20110214-optionsplus-v3.bas
Options File	icecleared_poweroptions_2012_05_21b.xlsx
Narrowed Slice Scenarios	O:\WC Risk\Common\stochastic-outcomes-current.wy.csv
Slice Scenarios	O:\WC Risk\Common\stochastic-outcomes-outer.wy.csv

Assets					
Thermal Parameters					
Plant Name	Operating HR	Fixed Startup Cost	Startup Gas (MMBTU)	Startup Power Output (MWh)	Capacity
8.9	2462	1272	134	50	

Plant Derates			
Plant Name	Begin Date	End Date	Derated Capacity
	4/9/2011	7/4/2011	0
	3/1/2012	4/1/2012	50

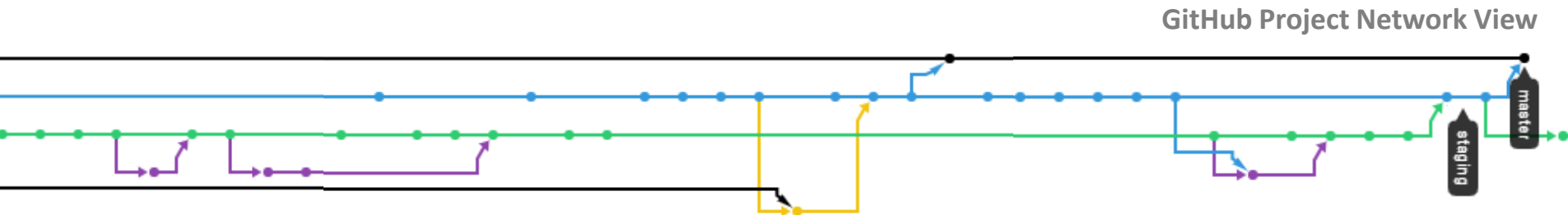
Retail Rates		
Begin Date	End Date	Average Rate
1/1/2010	12/31/2011	0
1/1/2012	12/31/2017	0

Variable Gen Parameters					
Plant Name	Type	Model	Capacity	Begin Date	End Date
	Wind	External			
	Wind	External			
	Hydro	External			
	Hydro	External			

CGS Derates		
Begin Date	End Date	Derated Capacity

Current Model

- Everything is in R.
 - The bulk of the model runs automatically overnight as a scheduled task.
 - Shiny UI enables on-demand analysis on the results.
- Differentiation between Dev, Staging, PROD environments.



Demo: Shiny App

Looks “Good” by Default

- Shiny templates are easy to use.
- No knowledge in HTML or JavaScript necessary
- Excel-looking interface possible by creating input tables and color-formatted data.

Excel Interface

		Date	Update Price & Record Greeks				Clear Manual Price Inputs				
		FPC Date									
		Heat Rate									
			42018	12019	22019	32019	42019	12020	22020	32020	42020
		Year	2018	2019	2019	2019	2019	2020	2020	2020	2020
		Quarter	4	1	2	3	4	1	2	3	4
9/12/18	Forward Price Curve	HLH	\$ 27.85	\$ 25.23	\$ 19.01	\$ 36.59	\$ 24.62	\$ 24.84	\$ 17.85	\$ 37.76	\$ 24.39
		LLH	\$ 23.53	\$ 20.70	\$ 10.19	\$ 22.46	\$ 21.17	\$ 20.84	\$ 11.03	\$ 22.84	\$ 20.03
		Gas	\$ 2.42	\$ 2.47	\$ 1.41	\$ 1.65	\$ 2.264	\$ 2.36	\$ 1.44	\$ 1.62	\$ 2.13
Manual Price Input		HLH									
		LLH									
		Gas									
Current Market		Spark - 7x24	\$ 3.40	\$ 0.27	\$ 0.12	\$ 13.48	\$ 1.67	\$ 0.98	\$ (0.38)	\$ 14.54	\$ 2.04
		Spark - HLH	\$ 5.29	\$ 2.25	\$ 3.97	\$ 19.73	\$ 3.20	\$ 2.73	\$ 2.59	\$ 21.13	\$ 3.97
		Spark - LLH	\$ 0.96	\$ (2.29)	\$ (4.85)	\$ 5.59	\$ (0.25)	\$ (1.28)	\$ (4.23)	\$ 6.21	\$ (0.39)
		Delta	0.60	0.50	0.49	1.00	0.72	0.61	0.30	1.03	0.75
		Gamma	0.13	0.15	0.25	0.00	0.18	0.11	0.14	0.01	0.16
Last Trade		Spark - 7x24	\$ 3.05	\$ 0.01	\$ -	\$ 5.76	\$ 1.36	\$ 1.75	\$ -	\$ 13.61	\$ -
		Spark - HLH	\$ 4.22	\$ 2.09	\$ -	\$ 9.62	\$ 3.44	\$ 3.60	\$ -	\$ 20.88	\$ -
		Spark - LLH	\$ 1.42	\$ (2.66)	\$ -	\$ 0.89	\$ (1.25)	\$ (0.65)	\$ -	\$ 4.43	\$ -
		Delta	0.95	0.53	32.75	0.95	0.84	0.64	-	1.05	-
		Delta Hedge Level	0.90	0.50	-	0.90	0.80	0.60	-	0.60	-
Trade Triggers		Gamma	0.03	0.19	-	0.03	0.18	0.10	-	0.01	-
		Buy/Sell	BUY	-	SELL	SELL	BUY	-	-	SELL	SELL
		HLH	25	0	-75	-25	0	0	0	-50	-100
		LLH	0	0	-50	0	25	0	0	-50	-75
		Gas	-2500	0	10000	2500	-2500	0	0	7500	15000

Shiny Interface

Monthly Trigger		On-Peak			Off-Peak								
		Oct-18	Nov-18	Dec-18	Jan-19	Feb-19	Mar-19	Apr-19	May-19	Jun-19	Jul-19	Aug-19	Sep-19
Quarter		4	4	4	1	1	1	2	2	2	3	3	3

Market

On-Peak Power Price	26.45	25.85	31.75	29.10	25.75	20.25	17.40	17.75	22.75	36.65	44.95	29.55
Off-Peak Power Price	23.10	22.40	25.30	25.70	20.40	16.75	9.80	9.30	11.40	19.75	26.55	20.70
Gas Price	2.01	2.18	2.99	2.94	2.44	1.90	1.38	1.33	1.32	1.77	1.74	1.72

Spark Spread

On-Peak	5.580	4.334	4.209	1.831	2.444	1.500	3.234	3.489	8.027	17.187	25.981	10.703
Off-Peak	2.234	0.891	-2.246	-1.555	-2.903	-2.009	-4.372	-4.953	-3.324	0.288	7.578	1.864

On-Peak Delta

Current Delta	1.000	1.000	1.000	0.829	0.874	0.811	0.226	0.302	1.000	1.000	1.000	1.000
Previous Delta												
Gamma	0.000	0.201	0.266	0.189	0.177	0.279	0.049	0.076	0.195	0.000	0.000	0.000

Off-Peak Delta

Current Delta	0.875	0.976	0.364	0.516	0.000	0.174	0.032	0.012	0.194	0.847	1.000	0.973
Previous Delta												
Gamma	0.045	0.283	0.150	0.245	0.293	0.218	0.039	0.011	0.089	0.073	0.005	0.012

Trade Triggers Hedge Unwind

On-Peak Power (MW)	0	0	0	25	-25	-75						
Off-Peak Power (MW)	0	0	0	0	0	0						
Gas (mmBtu/day)	0	0	0	-2500	2500	7500						

My Desktop & Workflow

WHAT DOES IT LOOK LIKE IN REALITY?

C:/Users/murphyey/Desktop/HedgeFox - staging - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Run_HedgeFox.r

Source on Save Run Source

```

13 ## author: Eina Ooka
14 ## created: May 2016
15 ##
16 #####
17
18 env.run <- "Dev"
19 market <- "CAISO"
20
21 ## =====
22 ## 0. Data & Enviro Prep
23 ## =====
24
25 # Read in market and run-environment from command argument
26 args = commandArgs(TRUE)
27 if(length(args) > 0){
28   market <- as.character(args[1])
29   env.run <- as.character(args[2])
30 }
31
32 # Configurations
33 config <- config::get(file = "config.yml", use_parent = FALSE
34                       , config = Sys.getenv("R_CONFIG_ACTIVE", env.
35 setwd(config$dir.main)
36
37 if(!tolower(Sys.getenv("USERNAME")) %in% tolower(unlist(strsplit(c
38   stop("Sorry. Ask the model maintainer to run the model for you.")
39
40 # Record time
41 ptm <- ptm.step <- utilitea::PrintRunLog(NULL, paste(market, "Run
42
43 <

```

Environment History Connections Git

Import Dataset

Global Environment

Environment is empty

Files Plots Packages Help Viewer

New Folder Delete Rename More

C: > Users > murphyey > Desktop > HedgeFox > RCode

	Name	Size	Modified
	..		
	.Rhistory	0 B	Mar 21, 2018, 1
	0.0_EnviroSetting.r	3.3 KB	Aug 27, 2018, 2
	0.1_StochVarSpecs.R	45 KB	Aug 27, 2018, 2
	0.2_Preliminary.r	3.2 KB	Aug 21, 2018, 1
	0.3_ScrapeData.r	5.7 KB	Aug 27, 2018, 2
	1.1_Decomposition.r	11.8 KB	Aug 27, 2018, 2
	1.2_LT.r	5.1 KB	Aug 21, 2018, 1
	1.3_STVolatility.r	3.2 KB	Aug 27, 2018, 2

Console C:/Users/murphyey/Desktop/HedgeFox/

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |

C:/Users/murphyeo/Desktop/HedgeFox - staging - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins HedgeFox

Run_HedgeFox.r

```

13 ## author: Eina Ooka
14 ## created: May 2016
15 ##
16 #####
17
18 env.run <- "Dev"
19 market <- "CAISO"
20
21 ## =====
22 ## 0. Data &
23 ## =====
24
25 # Read in ma
26 args = comma
27 if(length(ar
28   market <-
29   env.run <-
30 }
31
32 # Configurat
33 config <- co
34
35 setwd(config
36
37 if(!tolower
38   stop("Sorr
39
40 # Record tim
41 ptm <- ptm.
42
43 <

```

RStudio: Review Changes

Changes History staging Stage Revert Ignore

Commit message: Add Demo portfolio. |

Amend previous commit Commit

Show Staged Unstaged Context 5 line Ignore Whitespace Stage All Discard All

@@ -13,12 +13,12 @@

```

13 13 ## created: Nov 2016
14 14 ##
15 15 #####
16 16
17 17 env.run <- "dev"
18 18 market <- "PNW"
19 19 client.name <- " "
20 20
21 21
22 22 ## =====
23 23 ## 0. Data & Enviro Prep
24 24 ## =====

```

Console C:/Users/murphyeo

Type 'demo()' for some demos, help() for on-line help, or 'help.start()' for an HTML browser interface to help. Type 'q()' to quit R.

> |

Environment History Connections Git

Diff Commit

Staged Status Path

- .gitignore
- BasisForecast.Rproj
- Data/
- Diagnostic/
- Document/
- Model/
- RCode/ForecastBasis.r
- RCode/ForecastBasisModel.r
- Dev/BackcastAnalysis.r
- Dev/SQL/CreateTables.rmd

es Help Viewer

Delete Rename More

hyeo > Desktop > HedgeFox > RCode

	Size	Modified
	0 B	Mar 21, 2018
	3.3 KB	Aug 27, 2018
	45 KB	Aug 27, 2018
	3.2 KB	Aug 21, 2018
	5.7 KB	Aug 27, 2018
	11.8 KB	Aug 27, 2018
	5.1 KB	Aug 21, 2018
	3.2 KB	Aug 27, 2018

C:/Users/murphyeo/Desktop/HedgeFox - staging - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Environment History Connections Git

Global Environment

Environment is empty

```

24 ...
25
26 # Create Market Table
27
28 ```{sql, connection=conn}
29 CREATE TABLE Curve (
30   CurveID tinyint NOT NULL PRIMARY KEY,
31   Market varchar(10),
32   VarName varchar(50),
33   Commodity varchar(20),
34   Peak bit,
35   Forward bit,
36   BaseID tinyint,
37   RiskDBName varchar(50)
38 );
39 ...
40
41
42 ```{r}
43 df <- read.csv(glue("{dir.main}/RCode/dev/SQL/Curves.csv"))
44 dbwriteTable(conn, "Curve", df, overwrite = FALSE, append = TRUE,
45   row.names = FALSE)
46 ...
47
48 ```{r}
49 df <- dbgetquery(conn, glue("SELECT t1.*, t2.VarName AS Base FROM
50   Curve AS t1 LEFT OUTER JOIN Curve AS t2 ON t1.BaseID =
51   t2.CurveID;"))
52 df
53 ...
54
55 # Create Run Table
56
57 ```{sql, connection=conn}
58 CREATE TABLE Run (
59   RunID smallint NOT NULL PRIMARY KEY,
60   Market varchar(10),
61   ...
62 );
63 ...
64
65 # Create SQL Tables
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

```

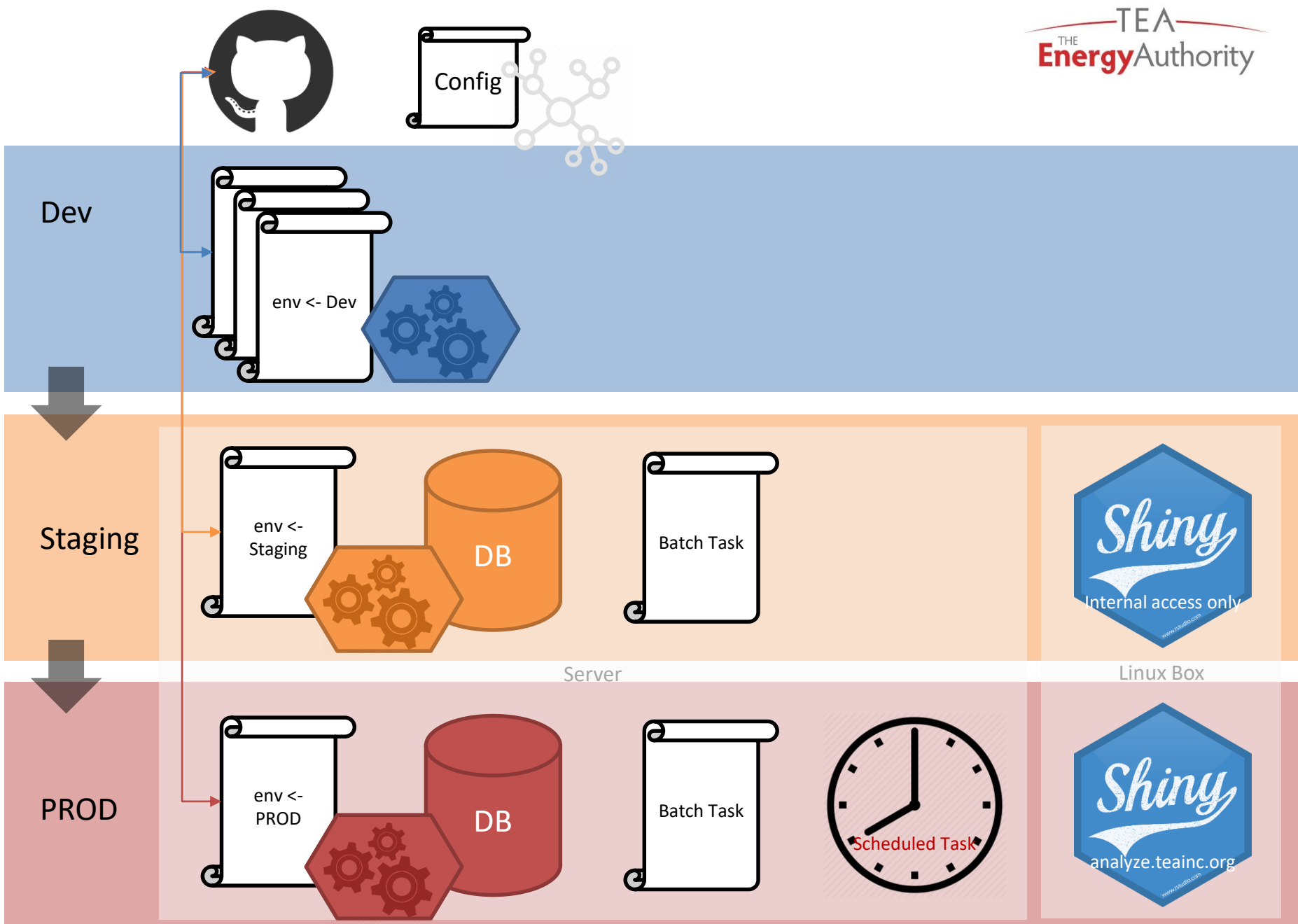
5:24 # Create SQL Tables R Markdown

Files Plots Packages Help Viewer

New Folder Delete Rename More

C: > Users > murphyeo > Desktop > HedgeFox > RCode

	Name	Size	Modified
	..		
	.Rhistory	0 B	Mar 21, 2018, 1
	0.0_EnviroSetting.r	3.3 KB	Aug 27, 2018, 2
	0.1_StochVarSpecs.R	45 KB	Aug 27, 2018, 2
	0.2_Preliminary.r	3.2 KB	Aug 21, 2018, 1
	0.3_ScrapeData.r	5.7 KB	Aug 27, 2018, 2
	1.1_Decomposition.r	11.8 KB	Aug 27, 2018, 2
	1.2_LT.r	5.1 KB	Aug 21, 2018, 1
	1.3_STVolatility.r	3.2 KB	Aug 27, 2018, 2



Take Away

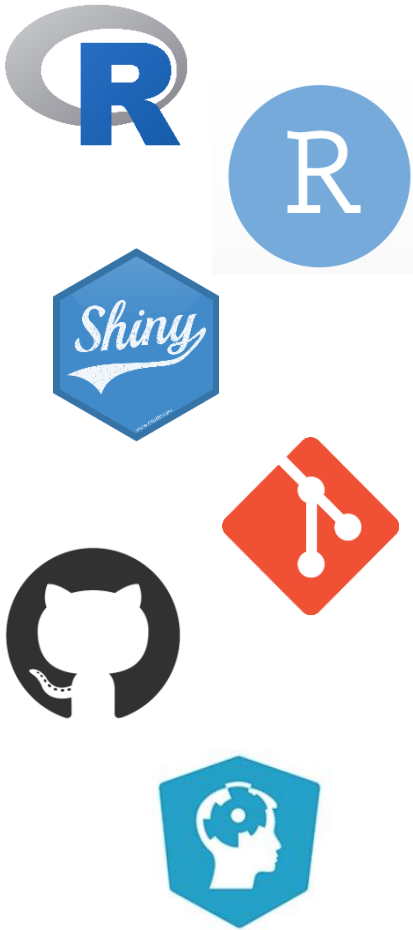
- Choosing the right set of platforms can help broaden our capability.
- We need multiple platforms that all work seamlessly together.
- Graduating from Excel takes a discipline, and not all projects need to move away from Excel. But if you choose to do so, have a vision before investing into a tool.

Thank you!

Contact:
eooka@teainc.org

APPENDIX

Resources



R

- R (<https://www.r-project.org/>)
- RStudio IDE (<https://www.rstudio.com/products/RStudio/>)

R Shiny

- R shiny (<https://shiny.rstudio.com/>)
- R Shiny Server (<https://www.rstudio.com/products/rstudio-server-pro/>)

Version Control

- Git (<https://git-scm.com/>)
- GitHub (<https://github.com/>)

Learn

- DataCamp (<https://www.datacamp.com>)

DataCamp Recommendations

For R Beginners

- Introduction to R
 - <https://www.datacamp.com/courses/free-introduction-to-r>
- Intermediate R
 - <https://www.datacamp.com/courses/intermediate-r>
- Data Visualization with ggplot2
 - <https://www.datacamp.com/courses/data-visualization-with-ggplot2-1>
- Building Web Applications in R with Shiny
 - <https://www.datacamp.com/courses/building-web-applications-in-r-with-shiny>
- Reporting with Rmarkdown
 - <https://www.datacamp.com/courses/reporting-with-r-markdown>
- Working with the RStudio IDE
 - <https://www.datacamp.com/courses/working-with-the-rstudio-ide-part-1>

