

Advocating for Automation: *Adapting Current Tools in Environmental Science through R*



Hannah Podzorski
rstudio::conf(2022)
July 27, 2022



Why Automate?



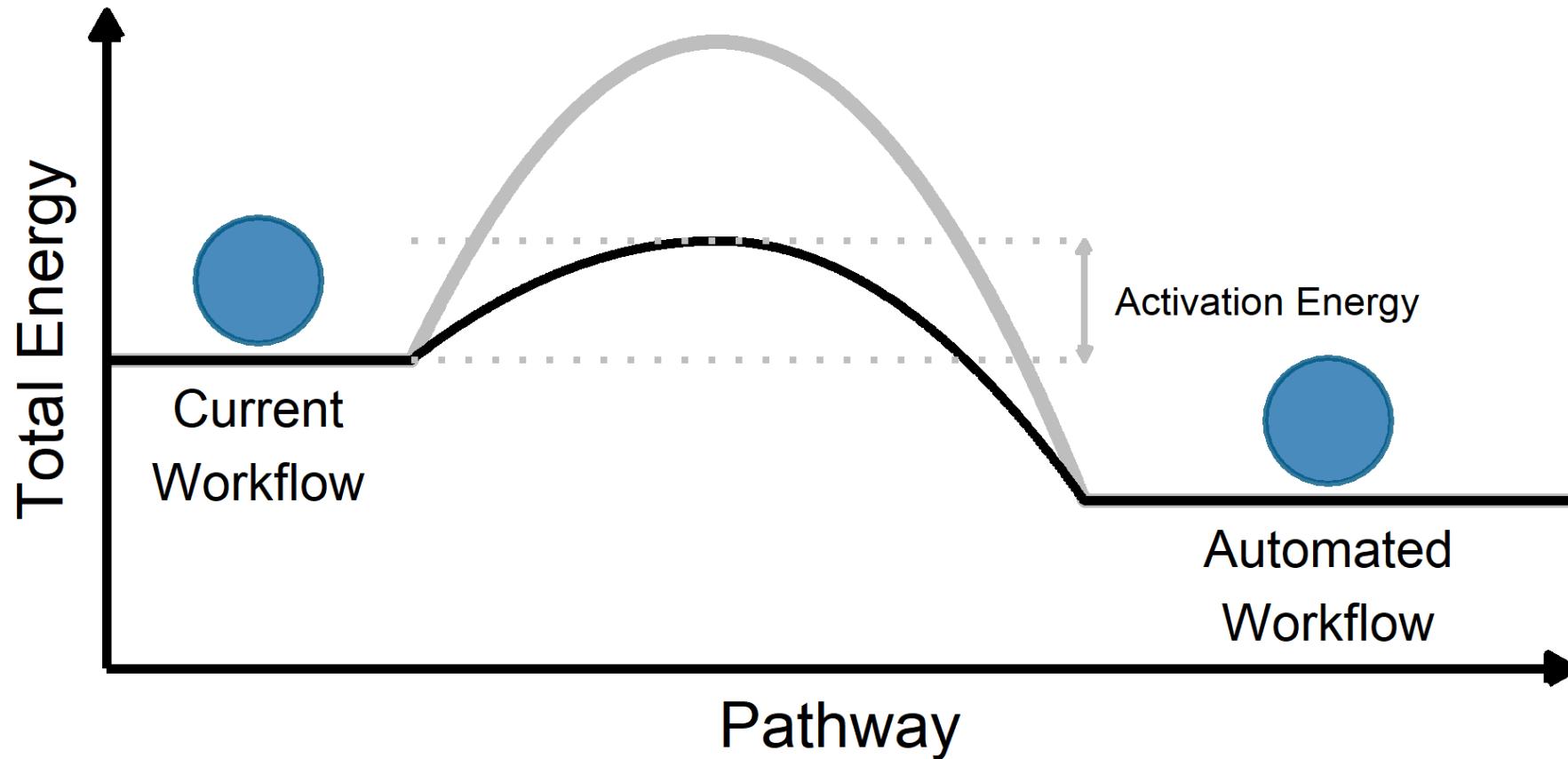
Programmer

**Task that takes
5 minutes**

Can it be automated?

<https://www.reddit.com/r/ProgrammerHumor/comments/f0ag3i/automation/>

Reducing the Activation Energy



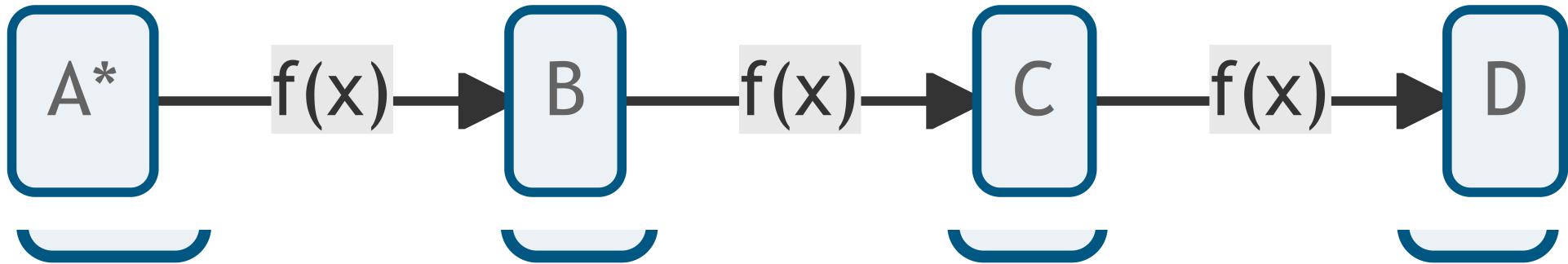
Let's Start with the Pitch

Differences in Workflow

Reactionary Workflow



Automated Workflow



Where to Start?

- Start small, task can be automated in the same amount of time as the original task.
- Meet team members where they are.

{openxlsx}

```
1 write.csv(data, "data.csv")
```

{openxlsx}

```
1 write.csv(data, "data.csv")
2
3 openxlsx::write.xlsx(data, "data.xlsx")
```

Well ID	Chemical Name	Units	Samples	First Date Sampled	Last Date Samples	Minium Sample	Mean Sample	Maximum Sample
MW-1	Calcium	mg/L	139	03/02/1999	12/15/2021	10	50.64	83
MW-1	Chloride	mg/L	138	03/02/1999	12/15/2021	17	89.68	150
MW-1	Fluoride	mg/L	138	03/02/1999	12/15/2021	0.12	0.33	0.63
MW-1	Magnesium	mg/L	138	03/02/1999	12/15/2021	3	16.04	26
MW-1	Potassium	mg/L	138	03/02/1999	12/15/2021	3.7	7.68	13
MW-1	Sodium	mg/L	138	03/02/1999	12/15/2021	27	75.83	120

Data from California's Groundwater Ambient Monitoring and Assessment Program (GAMA). Downloaded 2022-07-11.

{officer}

```
1 library(officer)
2
3 plot <- rvg::dml(ggobj = plot)
4
5 ppxx <-read_pptx() %>%
6   add_slide() %>%
7   ph_with(plot, ph_location(left = 1.3, top = 0.4, width = 8.75, height = 6.9))
8
9 print(ppxx, "./R/Fig-Example.pptx")
```

{officer}



Delayed Gratification

ProUCL

- Statistical Software for Left Censored Environmental Data
 - Calculates Upper Confidence Limits (UCLs)
- Developed by the U.S Environmental Protection Agency (EPA)



ProUCL Automation

ProUCL Output

Magnesium_MG/L (100832)

General Statistics

Total Number of Observations	137	Number of Distinct Observations	38
		Number of Missing Observations	1
Minimum	3	Mean	16.07
Maximum	26	Median	18
SD	5.023	Std. Error of Mean	0.429
Coefficient of Variation	0.313	Skewness	-0.783

Normal GOF Test

Shapiro Wilk Test Statistic	0.914	Shapiro Wilk GOF Test
5% Shapiro Wilk P Value	1.138E-10	Data Not Normal at 5% Significance Level
Lilliefors Test Statistic	0.165	Lilliefors GOF Test
5% Lilliefors Critical Value	0.0761	Data Not Normal at 5% Significance Level

Data Not Normal at 5% Significance Level

Assuming Normal Distribution

95% Normal UCL	95% UCLs (Adjusted for Skewness)
95% Student's-t UCL	16.78
	95% Adjusted-CLT UCL (Chen-1995) 16.74 95% Modified-t UCL (Johnson-1978) 16.78

Final Thoughts

- It's ok to start small.
- All skill sets welcome!
- Focus on communication.

Questions?

GitHub Repo:
*github.com/
hannahpodzorski/
advocating-for-
automation*



Contact Information:
hpodzorski@gsi-net.com
twitter - @hpodz
github - hannahpodzorski

Agenda

- Developing a ‘Pitch’
- Where to start?
- Unconventional Automation (Case Study)

DISCLAIMER