Using R to Analyze Recruiting Pipelines

Maryam Jahanshahi
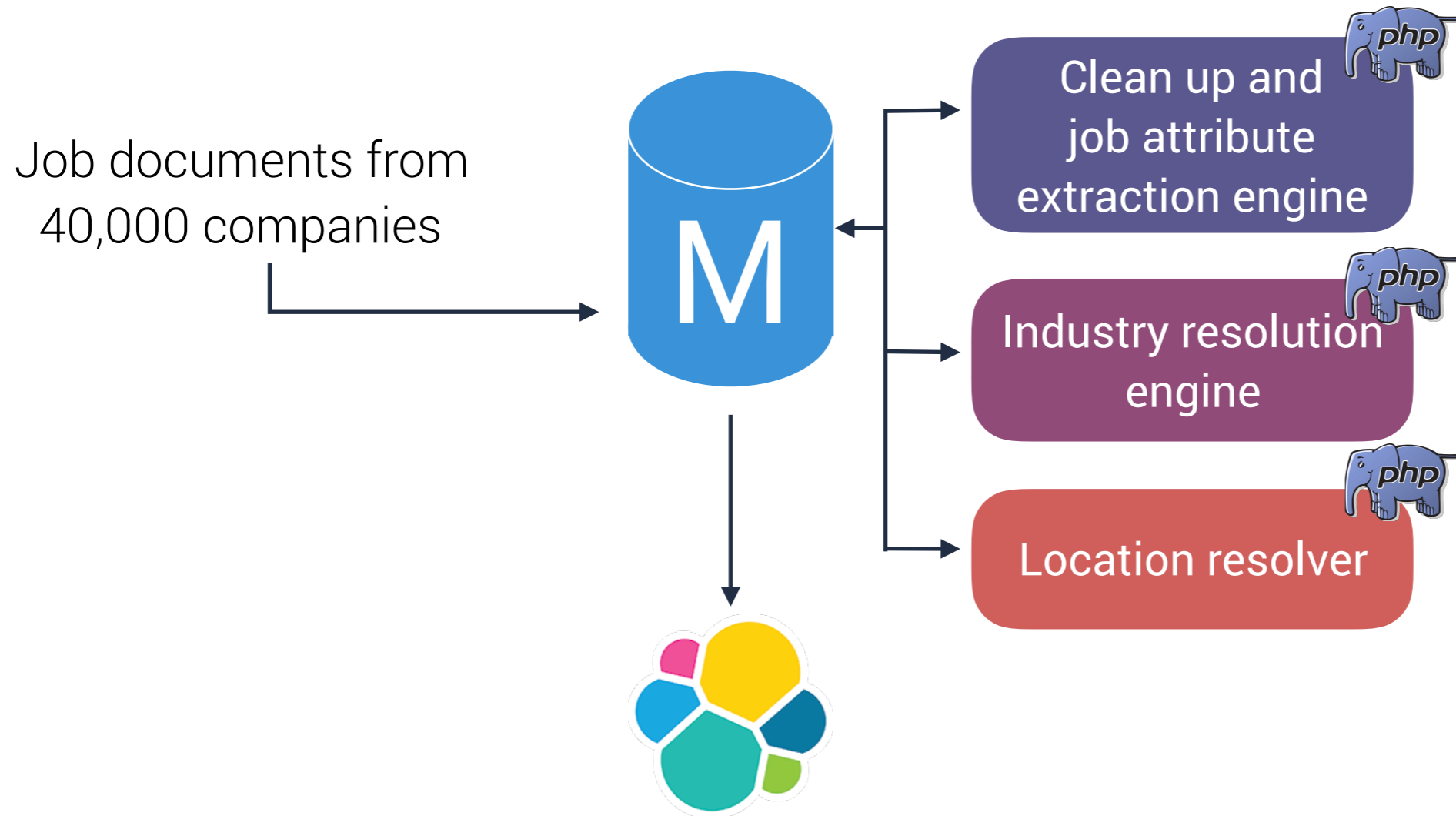Scientist @ TapRecruit
@mjahanshahi

tapRecruit

Jenny Dearborn @DearbornJenny

If you were to ask your CFO to make a decision, she would rely on facts, not opinions.
In contrast, for years, the HR function has been asked to resolve some of the thorniest problems in the organization **without the clarity of meaningful data analytics.**

The TapRecruit Hierarchy of Recruiting Needs

Time-to-fill is slow

>120 days for Tier 1 City

Candidate pool is not diverse

Few applications from women and POC

Wrong type of candidates

A mid-level role is only attracting fresh grad apps
Hiring team wants finance exp. but attracting programmers

Not enough applications

Junior Jobs: <30 apps in Tier 1 City
Mid-Level Jobs: <20 apps in Tier 1 City

@mjahanshahi

Transforming unstructured corpuses to structured taxonomies

Job documents from 40,000 companies

M

Clean up and job attribute extraction engine

Industry resolution engine

Location resolver

@mjahanshahi

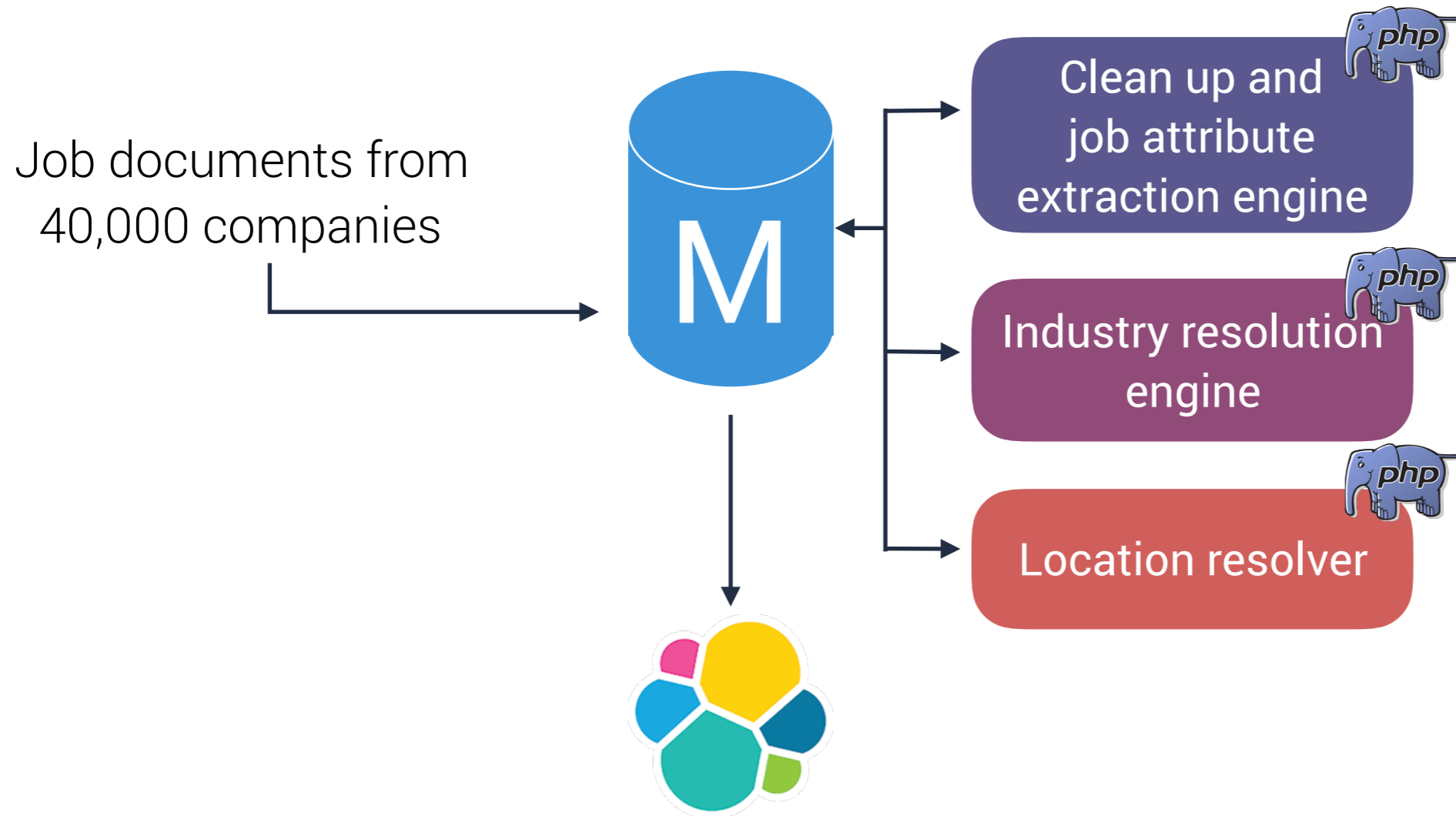# Transforming unstructured corpuses to structured taxonomies

Data Scientist

Hooli.xyz - Hooli HQ

Hooli is looking for a data scientist to join a team passionate about Marketing Analytics for our suite of media products. You will work with internet-scale data across numerous customer touch points, developing capabilities tied to audience...
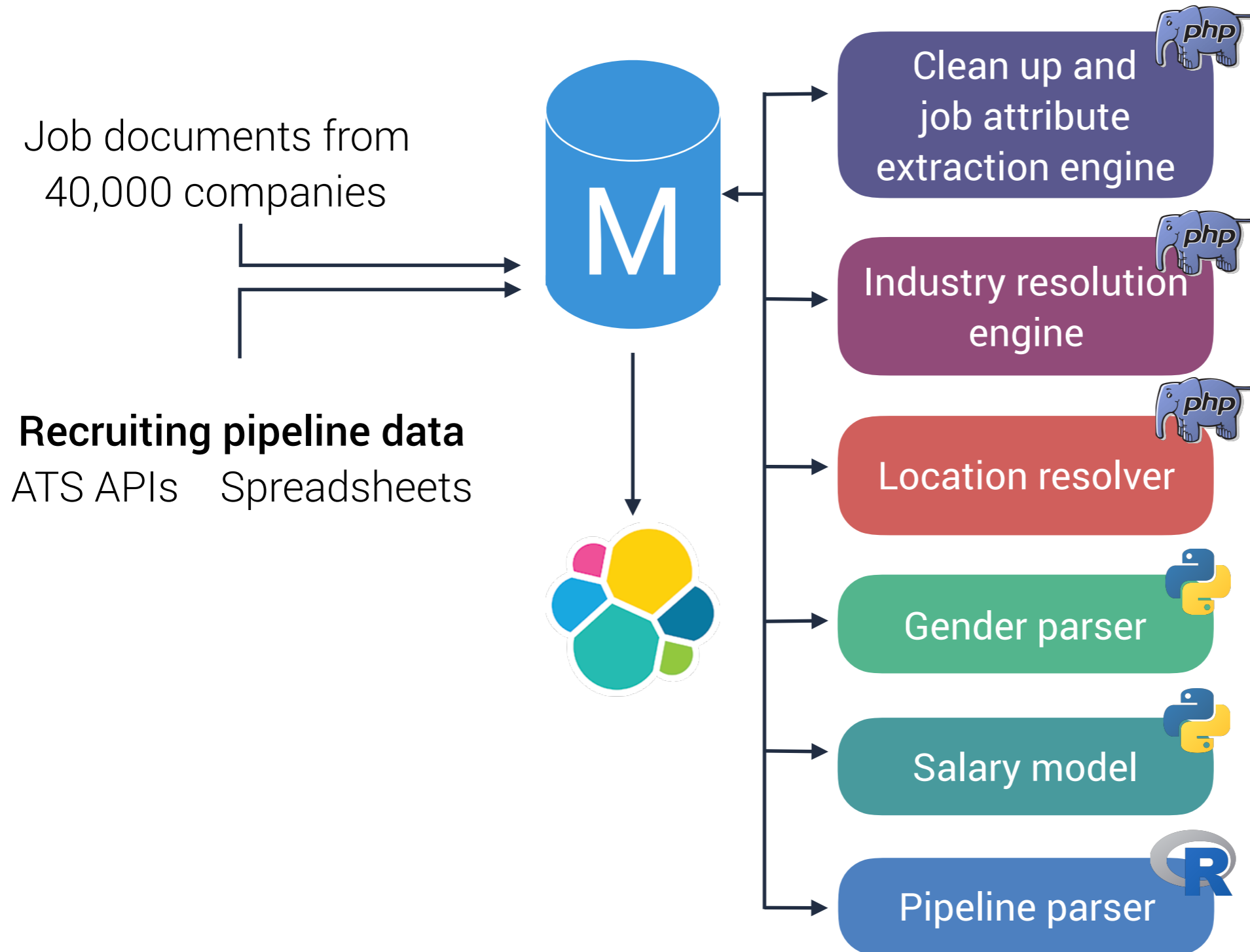
M

Clean up and job attribute extraction engine

Industry resolution engine

Location resolver
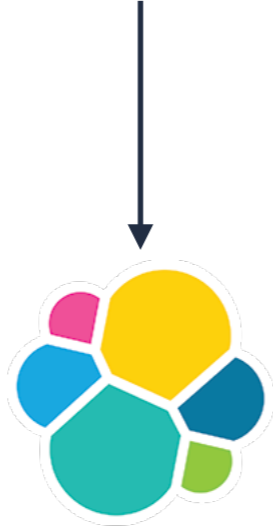
Junior

Programming
Data & Analytics
Data Visualization

BigTech
Media Technology
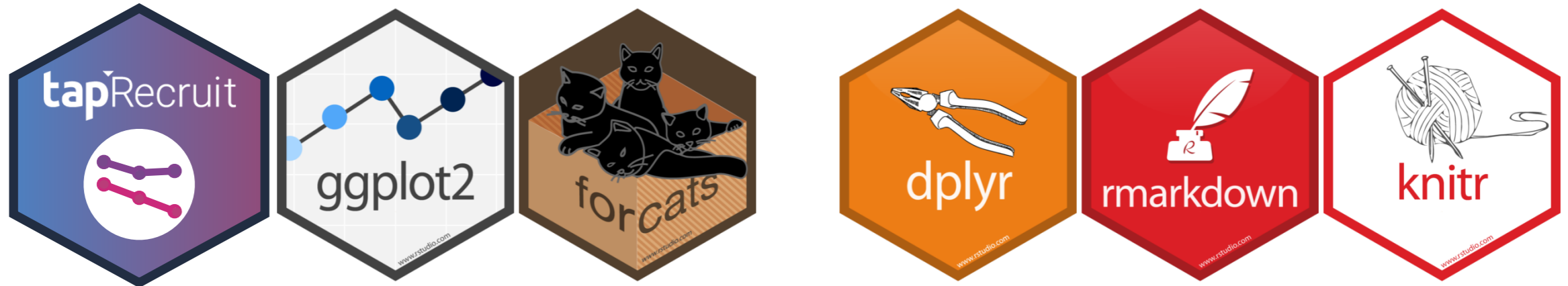Fortune500

Mountainview, CA

Python
Hadoop
SQL

R
Tableau

@mjahanshahi

Transforming unstructured corpuses to structured taxonomies

Job documents from 40,000 companies

M

Clean up and job attribute extraction engine

Industry resolution engine

Location resolver

Pipeline Analytics involves deep integration of multiple resolution engines

Job documents from 40,000 companies

Recruiting pipeline data
ATS APIs   Spreadsheets

Clean up and job attribute extraction engine

Industry resolution engine

Location resolver

Gender parser

Salary model

Pipeline parser

@mjahanshahi

Making Recruiting Reproducible (Again?) requires Reproducible Workflows

Data cleanup scripts
- Integrity checking
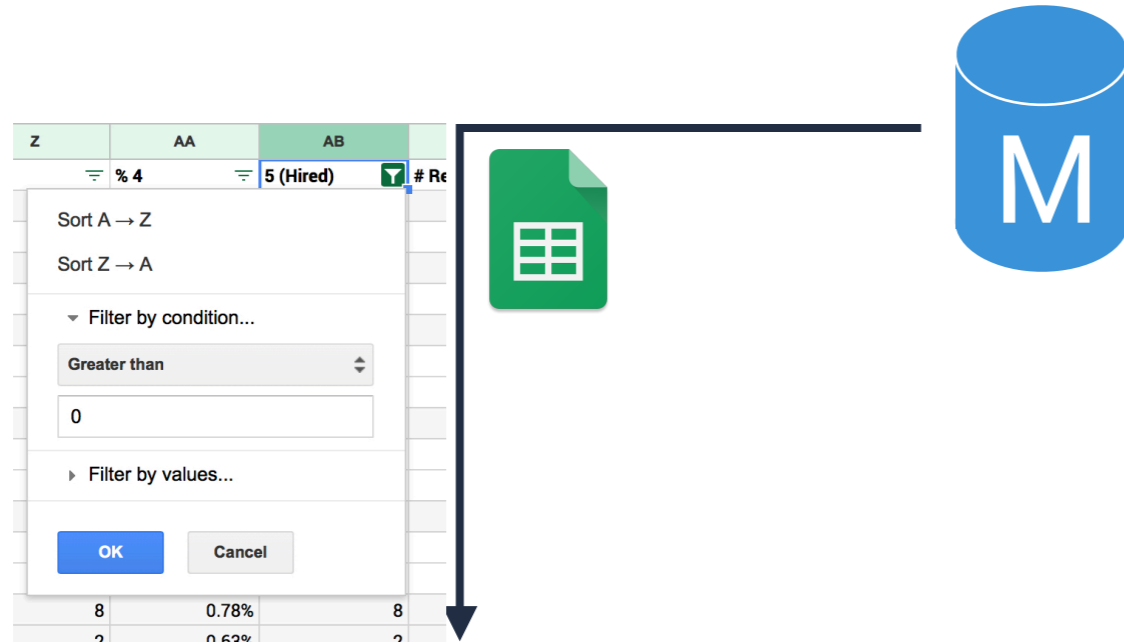
Basic data transformation
- Appropriate factor levels

Graph templates

R Markdown notebooks
- Visualizing distributions
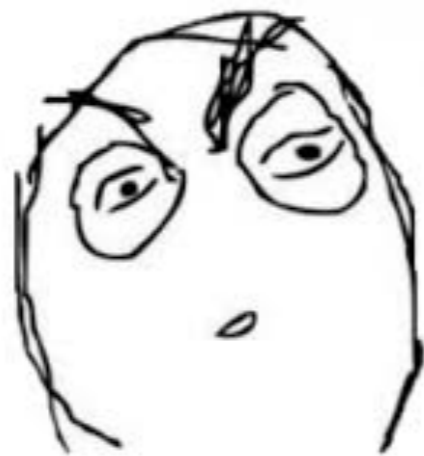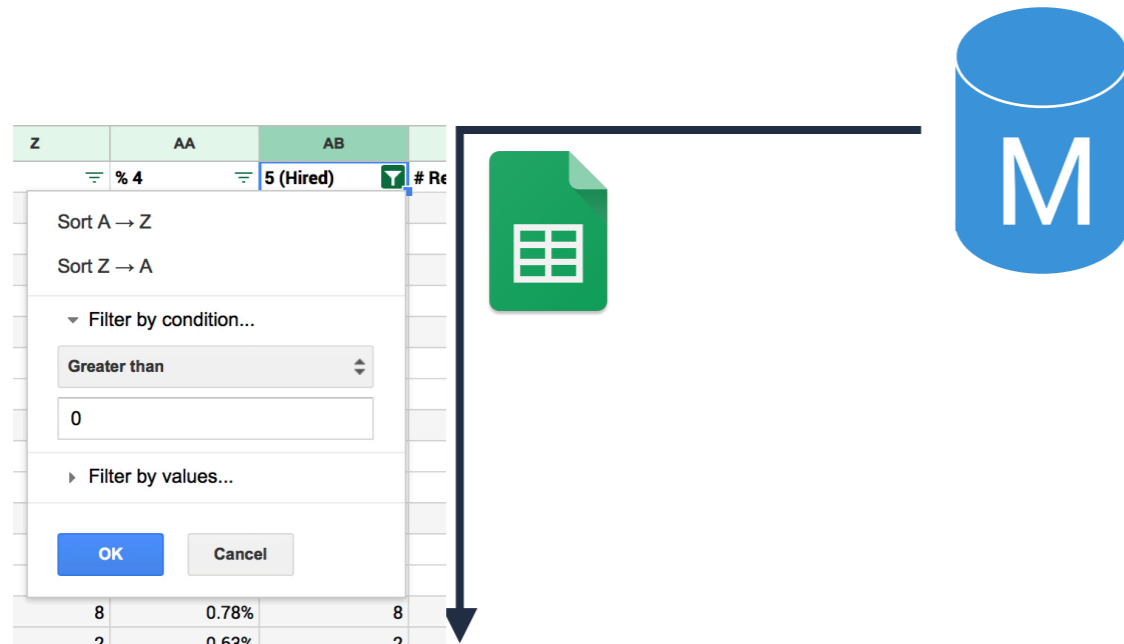- Company-specific data transformations

Graphs for reporting

@mjahanshahi

Making Recruiting Reproducible (Again?) requires Reproducible Workflows
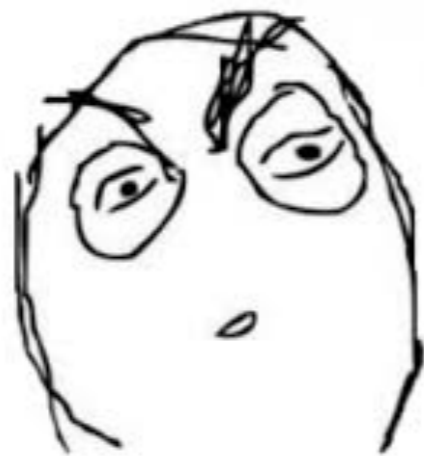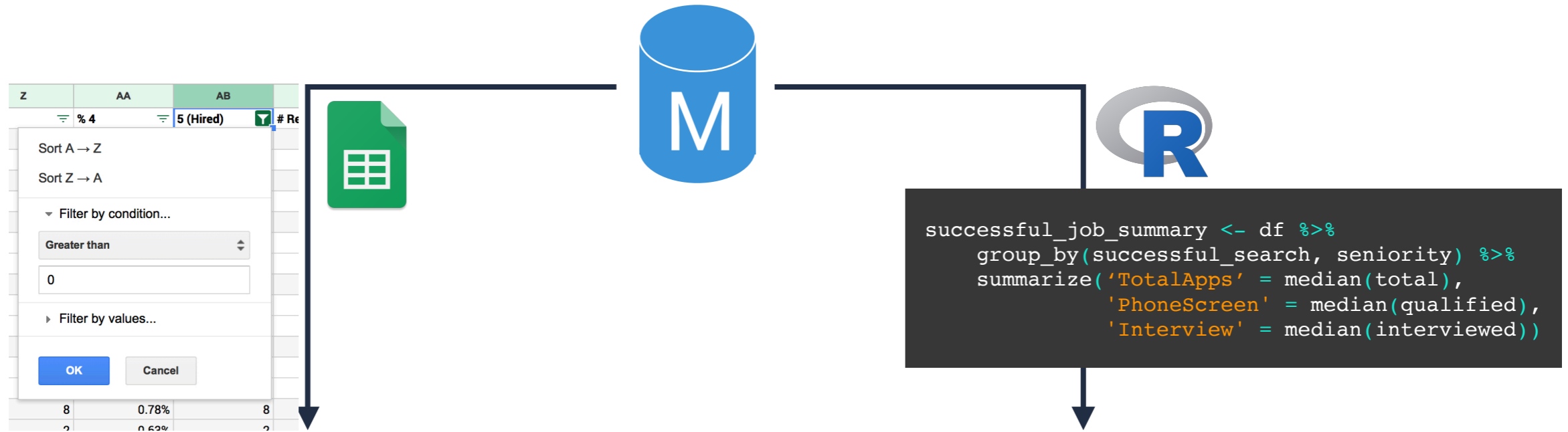
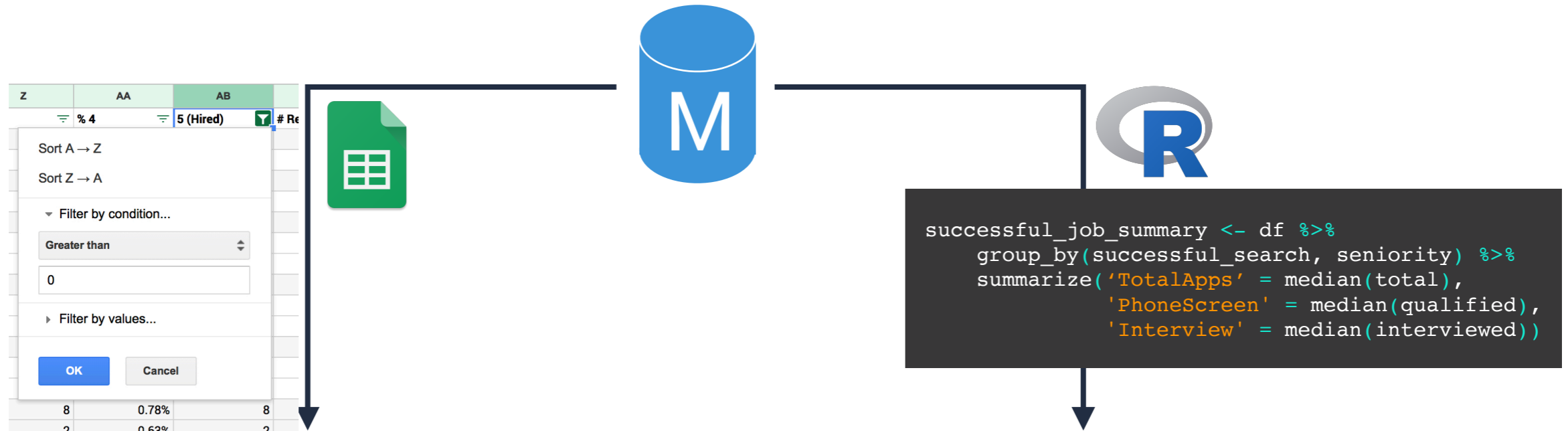# Making Recruiting Reproducible (Again?) requires Reproducible Workflows

Making Recruiting Reproducible (Again?) requires Reproducible Workflows


```
successful_job_summary <- df %>%
    group_by(successful_search, seniority) %>%
    summarize('TotalApps' = median(total),
              'PhoneScreen' = median(qualified),
              'Interview' = median(interviewed))
```
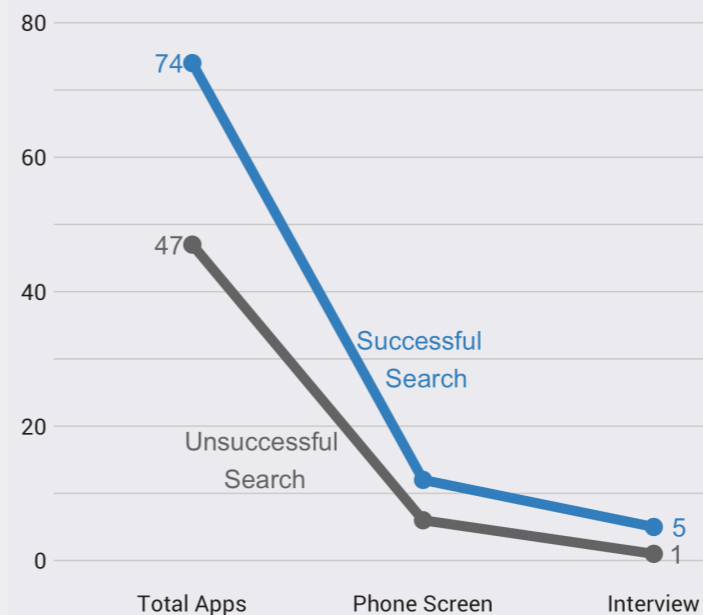
Why you no work?!

Making Recruiting Reproducible (Again?)
requires Reproducible Workflows


```
successful_job_summary <- df %>%
    group_by(successful_search, seniority) %>%
    summarize('TotalApps' = median(total),
              'PhoneScreen' = median(qualified),
              'Interview' = median(interviewed))
```

Applicants by Stage

Jobs that do not result in an offer start with fewer candidates at the start of the funnel.

Female Applicants by Stage

Jobs without offers to female applicants show female representation issues throughout the funnel.