

Welcome to the course!

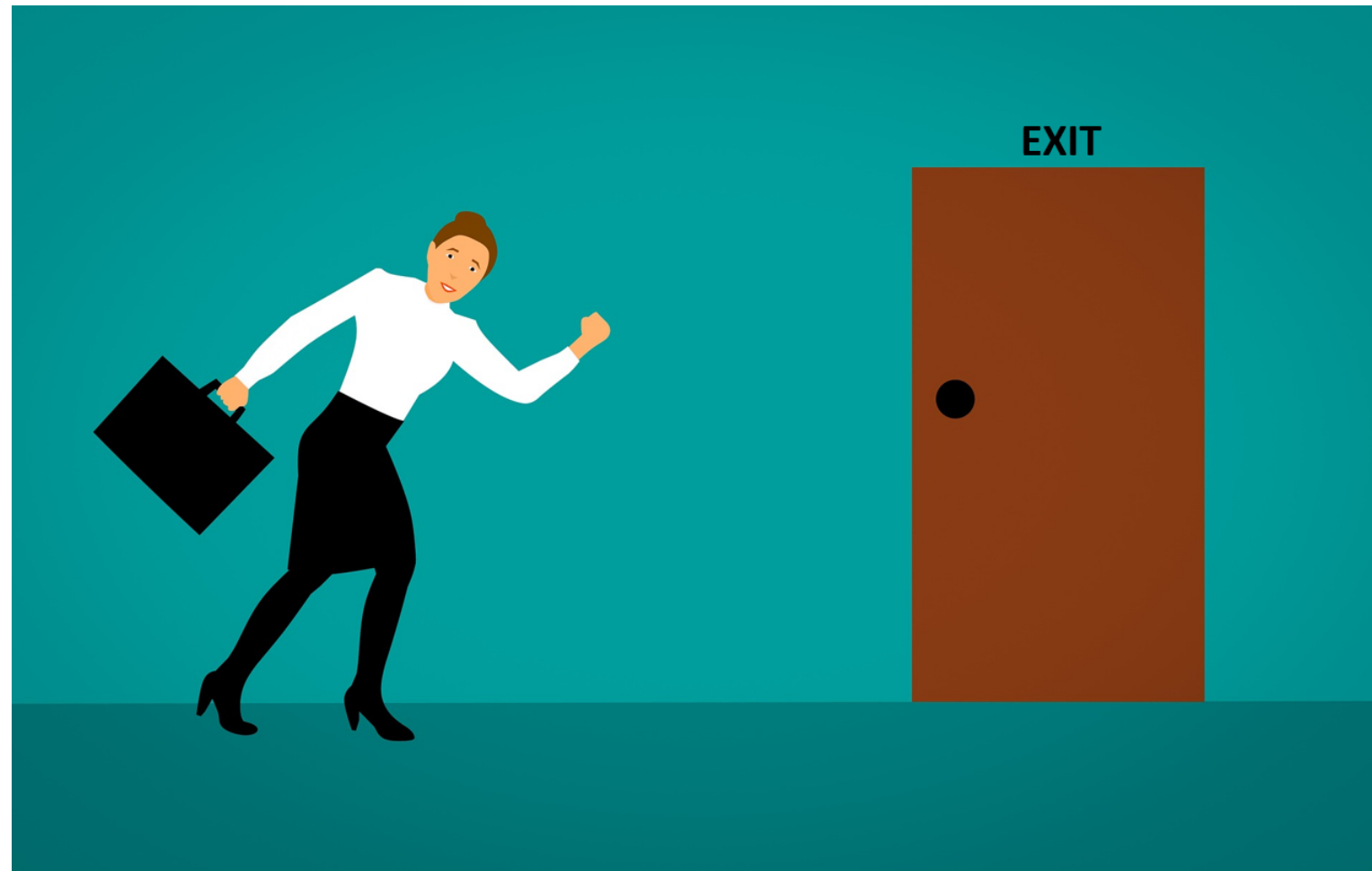
HUMAN RESOURCES ANALYTICS: PREDICTING EMPLOYEE CHURN IN R



Anurag Gupta and Abhishek Treh...
People Analytics Practitioners

Understanding employee turnover

- Churn refers to the gradual loss of employees over a period of time
- **Churn / Turnover / Attrition** are interchangeably used



Why employee turnover matters?

- Employee turnover is the biggest issue facing HR
- Employee turnover is the highest its been in 10 years
- Turnover costs way more than you think

Types of employee turnover

VOLUNTARY TURNOVER

- When an employee chooses to resign

INVOLUNTARY TURNOVER

- When an organization decides to let go of an employee

Common reasons for employee turnover

- Better opportunity
- Health
- Relocation
- Education
- Personal reasons etc.

Hidden reasons of employees turnover

- Relationship with manager
- Percent salary hike
- Overtime
- Travel distance
- Career satisfaction
- Tenure

Course overview

- Chapter 1: Introduction to employee turnover prediction
- Chapter 2: Building relationship with data
- Chapter 3: Building turnover prediction model using logistic regression
- Chapter 4: Model validation, RoI calculation and retention strategy

Basic requirements for the course

- dplyr: Data wrangling
- ggplot2: Visualization and exploration

Let's practice!

HUMAN RESOURCES ANALYTICS: PREDICTING EMPLOYEE CHURN IN R

Know more about turnover

HUMAN RESOURCES ANALYTICS: PREDICTING EMPLOYEE CHURN IN R



Anurag Gupta

People Analytics Practitioner

Understanding the data

```
glimpse(org)
```

```
Observations: 2,291
Variables: 12
$ emp_id      <chr> "E11061", "E1031", "E6213", "E5900", "E3044"...
$ status      <chr> "Inactive", "Inactive", "Inactive", "Inactiv...
$ turnover    <int> 1, 1, 1, 1, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 1,...
$ location    <chr> "New York", "New York", "New York", "New Yor...
$ level       <chr> "Analyst", "Analyst", "Analyst", "Analyst", ...
$ date_of_joining <chr> "22-03-2012", "09-03-2012", "06-01-2012", "2...
$ last_working_date <chr> "11-09-2014", "05-06-2014", "30-04-2014", "0...
$ gender      <chr> "Male", "Female", "Female", "Female", "Femal...
$ department  <chr> "Customer Operations", "Customer Operations"...
$ mgr_id      <chr> "E1712", "E10524", "E4443", "E3638", "E3312"...
$ cutoff_date <chr> "31-12-2014", "31-12-2014", "31-12-2014", "3...
$ emp_age     <dbl> 22.49, 22.42, 22.24, 22.32, 22.14, 22.67, 22...
```

Calculating turnover rate

$$\text{Turnover rate} = \frac{\text{Number of employees who left}}{\text{Total number of employees}}$$

or

$$\text{Turnover rate} = \frac{\text{Count of all 1's}}{\text{Count of all 1's} + \text{Count of all 0's}} = \text{mean}(\text{turnover})$$

where

1 means Inactive ; 0 means Active

Count Active and Inactive employees

```
# Count Active and Inactive employees
```

```
org %>%  
  count(status)
```

```
# A tibble: 2 x 2  
  status      n  
  <fct>    <int>  
1 Active   1881  
2 Inactive  410
```

Calculate turnover rate

```
# Calculate average turnover rate

org %>%
  summarize(turnover_rate = mean(turnover))
```

```
  turnover_rate
1      0.1789612
```

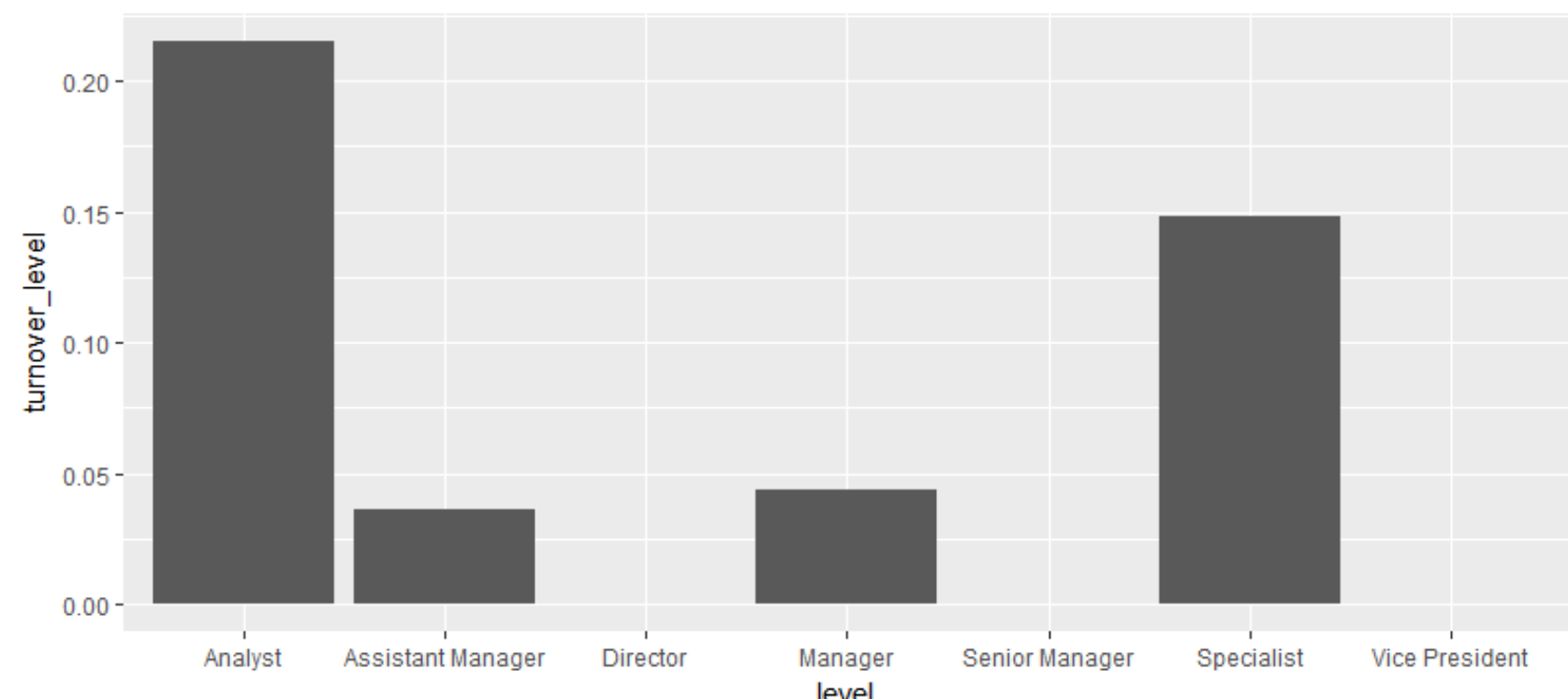
Calculate turnover rate at each level

```
df_level <- org %>%  
  group_by(level) %>%  
  summarize(turnover_level = mean(turnover))  
df_level
```

```
# A tibble: 7 x 2  
  level      turnover_level  
  <fct>      <dbl>  
1 Analyst      0.215  
2 Assistant Manager 0.0365  
3 Director      0  
4 Manager      0.0435  
5 Senior Manager  0  
6 Specialist    0.149  
7 Vice President  0
```

Visualize the turnover trends using ggplot

```
# Visualize the results
library(ggplot2)
ggplot(df_level, aes(x = level, y = turnover_level)) +
  geom_col()
```



Let's practice!

HUMAN RESOURCES ANALYTICS: PREDICTING EMPLOYEE CHURN IN R

Talent segments

HUMAN RESOURCES ANALYTICS: PREDICTING EMPLOYEE CHURN IN R



Anurag Gupta

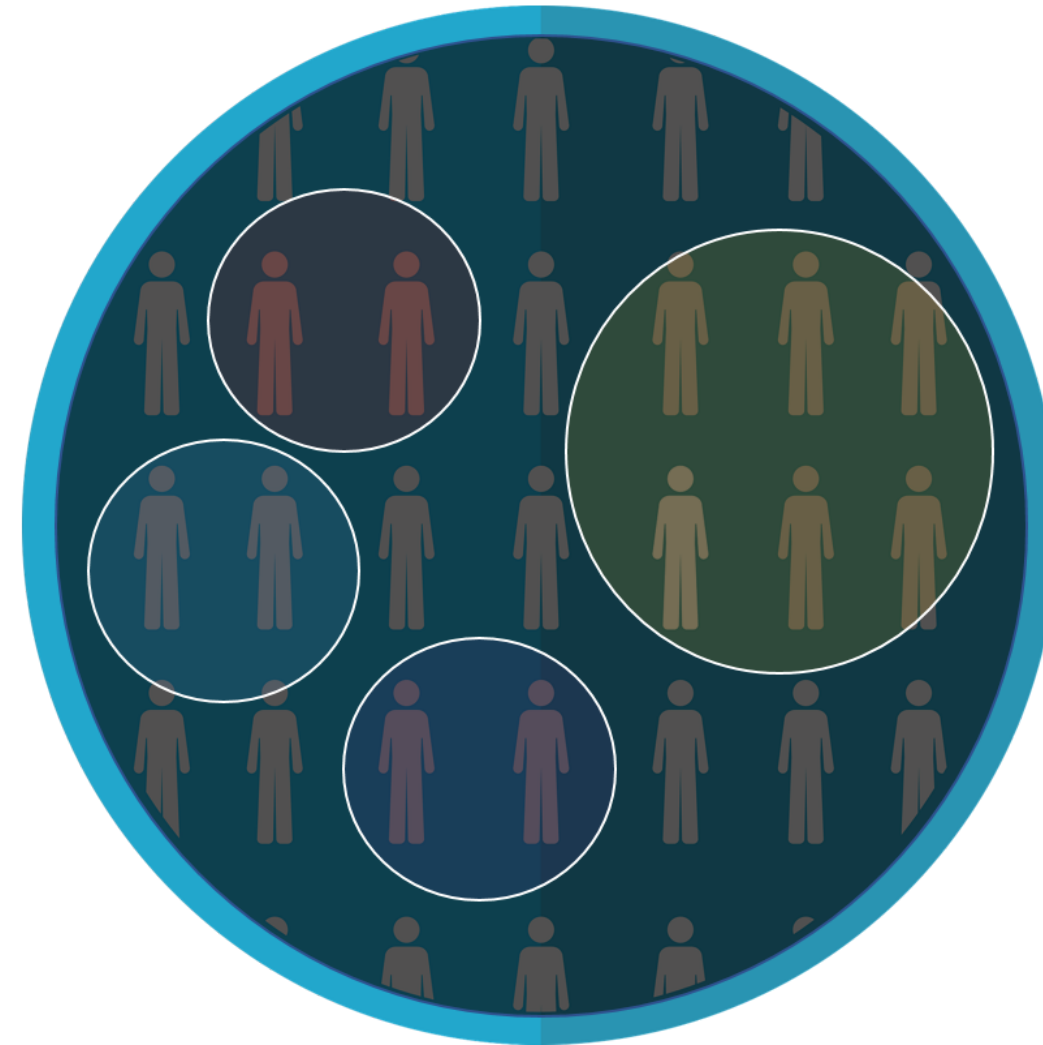
People Analytics Practitioner

Identifying the talent segments

Entire Population

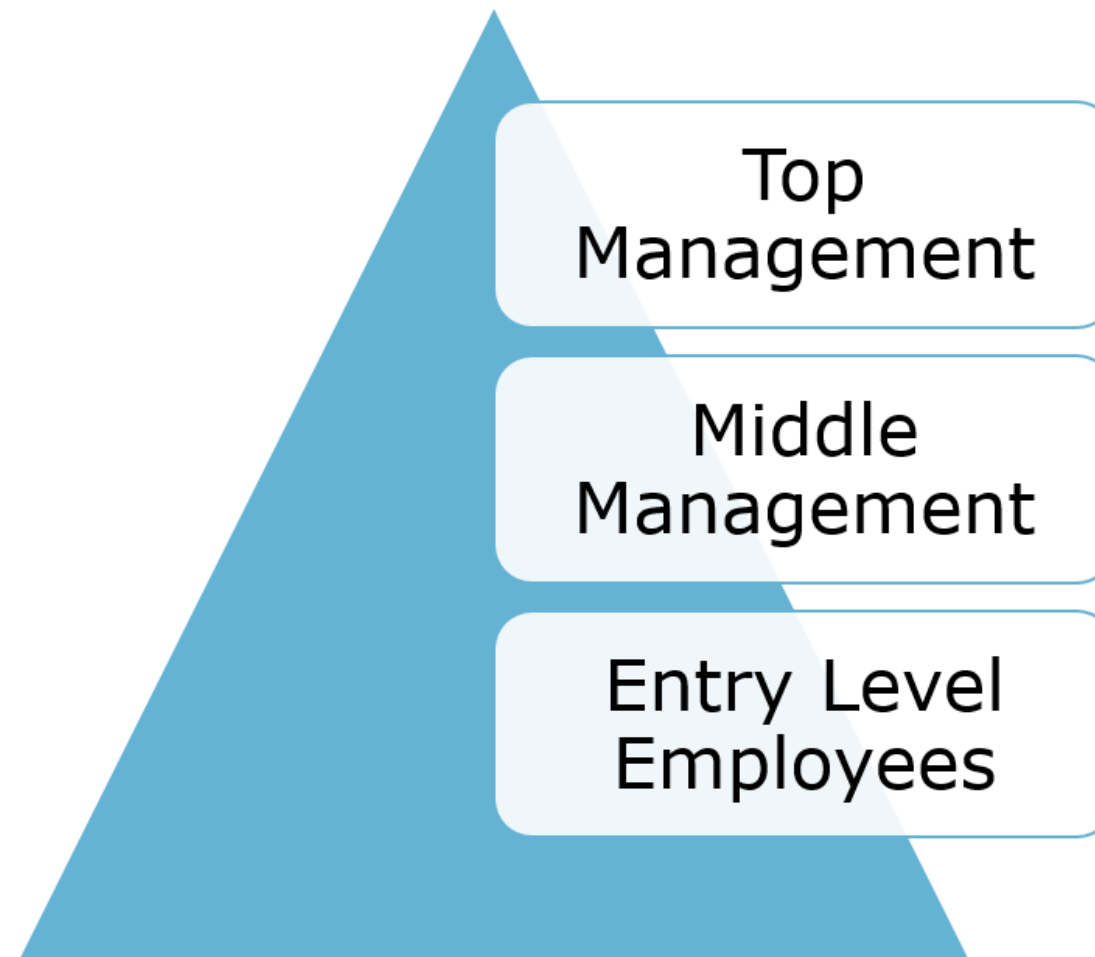


Talent Segments



Identifying the talent segments

Organization



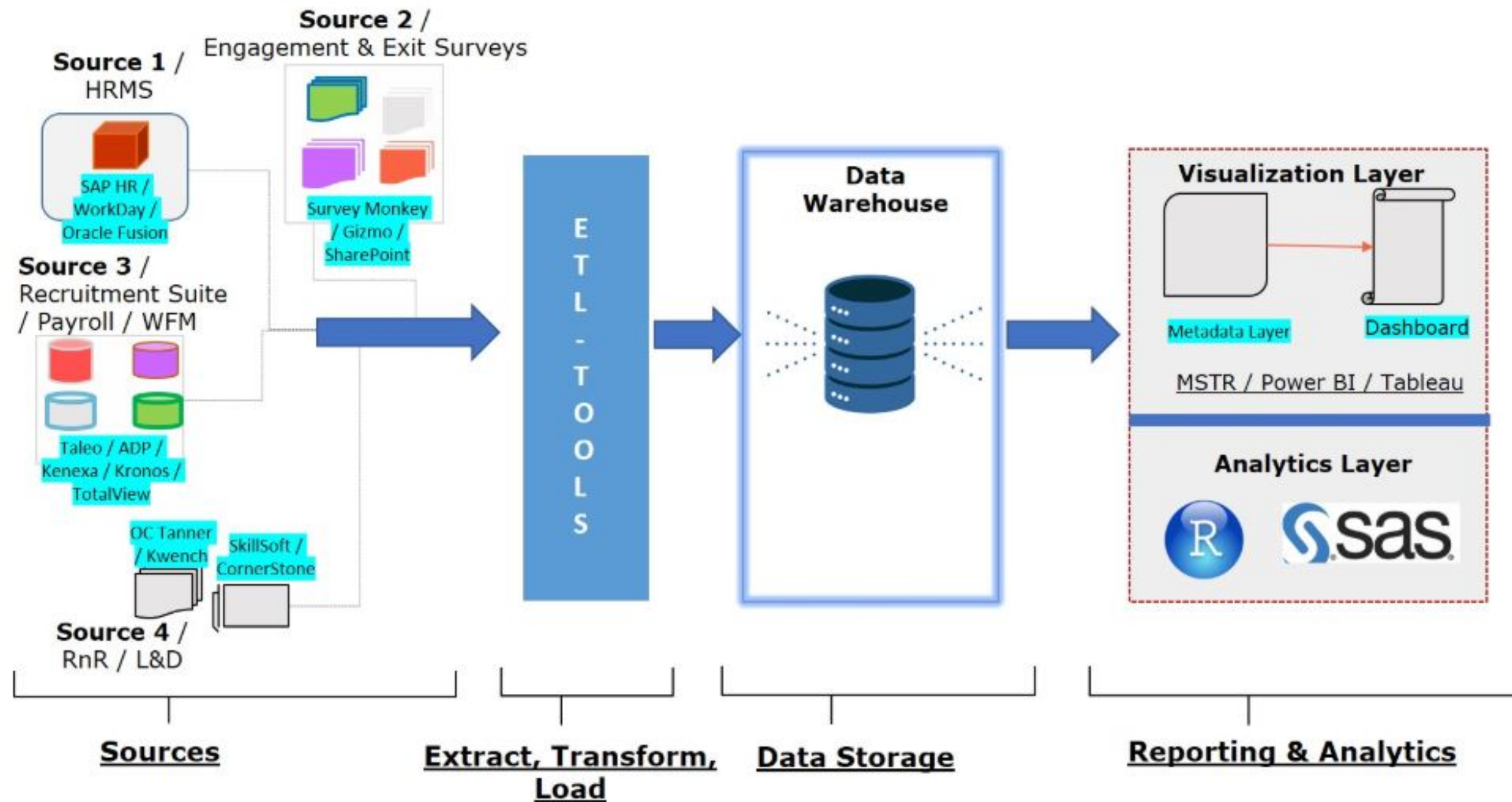
Filtering the dataset

```
# Select the employees at Analyst and Specialist level  
org2 <- org %>%  
  filter(level %in% c("Analyst", "Specialist"))
```

HR data sources across employee life cycle (ELC)

Employee Life Cycle	HR Data Source
Talent Acquisition	Taleo, ADP
Employee Info	SAP HR, Workday, SuccessFactors
Learning & Development	SkillSoft, Cornerstone
Rewards & Recognition	OC Tanner, Kwench
Onboarding, Engagement and Exit Surveys	SurveyMonkey, SurveyGizmo

HR data architecture



Merge datasets using left_join()

```
glimpse(df1)
```

```
Observations: 2  
Variables: 2  
$ emp_id <int> 1, 2  
$ level  <fct> Analyst, Analyst
```

```
glimpse(df2)
```

```
Observations: 2  
Variables: 2  
$ emp_id <int> 1, 2, 3  
$ location <fct> NYC, Atlanta, NYC
```


Let's practice!

HUMAN RESOURCES ANALYTICS: PREDICTING EMPLOYEE CHURN IN R