

# Leveling up

INTRODUCTION TO SPARK WITH SPARKLYR IN R



**Richie Cotton**

Data Evangelist at DataCamp

```
abba <- data_frame(  
  agnetha    = 1:3,  
  benny      = pi ^ c(1, 4, 9),  
  bjorn      = month.name[1:3],  
  anni_frid  = c(TRUE, FALSE, TRUE)  
)  
abba %>%  
  select(starts_with("a"))
```

```
# A tibble: 3 x 2  
  agnetha anni_frid  
    <int>   <lgl>  
1       1     TRUE  
2       2    FALSE  
3       3     TRUE
```

# Computing and collecting

```
intermediate_data <- initial_data %>%  
  # some calculations  
  compute("an_intermediate_result")  
  
results <- intermediate_data %>%  
  # some more calculations  
  collect()
```

# SQL and database joins

```
sql_query <- "SELECT agnetha, bjorn FROM abba where anni_frid"  
results <- dbGetQuery(spark_conn, sql_query)
```

```
  agnetha  bjorn  
1         1 January  
2         3  March
```

# Let's practice!

INTRODUCTION TO SPARK WITH SPARKLYR IN R