# Flexible Class Session #1

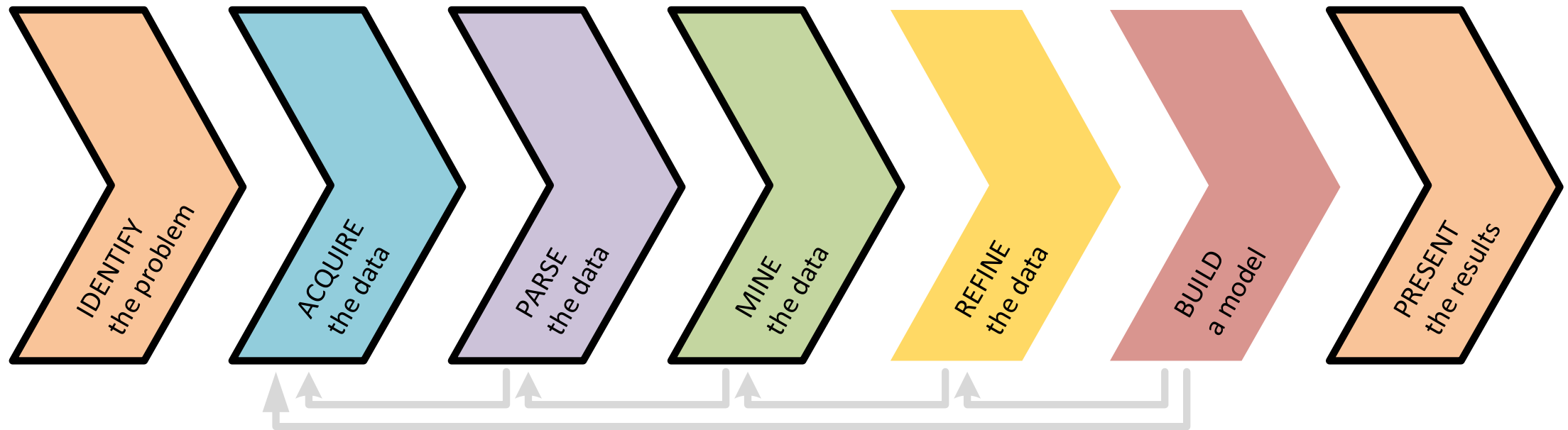*Ivan Corneillet*

*Data Scientist*

# Learning Objectives

‣ Lab session (no lecture today) to put into practice Unit 1's concepts:

  ‣ Research Design

  ‣ Data Manipulation, dataset tidying and Exploratory Data Analysis with *pandas*

  ‣ Statistics and Visualization with *pandas*

‣ Also, an opportunity to collaborate as a team using git/GitHub

‣ Have Fun doing Data Science work!

# Outline

‣ Lab

‣ Presentations and Q & A

‣ In-flight

   ‣ **Unit Project 2 (due next session on ~~3/10~~ 3/15)**

   ‣ Final Project 1 (due in 1.5 weeks)

# Today's Lab and the Data Science Workflow



IDENTIFY the problem

ACQUIRE the data

PARSE the data

MINE the data

REFINE the data

BUILD a model

PRESENT the results

# Lab
# Storms, *pandas*, and GitHub

# Q & A

# Before Next Class

# Before Next Class

‣ Understand the difference between vectors, matrices, *pandas Series*, and *pandas DataFrames*

‣ Understand the concepts of outliers and distance

‣ Effectively show correlations between an independent variable $X$ and a dependent variable $Y$

‣ Be able to interpret t-values, p-values, and confidence intervals

# Exit Ticket

*Don't forget to fill out your exit ticket* *here*