Capstone Final Project

Guillaume Babo

7/13/2017

## What is the purpose of the project?

In this project, my goal is to find correlations between the causes of the oil spill and the pipeline type (aboveground, underground or tank) and then how they relate to the different cost associated with the spill. More precisely,the questions i will answer are : - The most common causes of spills most likely happen overground or underground? - What are the correlations between causes and costs? \_ What are the correlations between pipeline type and costs?

## Impact on client decision

The targeted clientele would be pipeline companies (design or construction). Based on this study the client would definetely be able to adjust their design startegy knowing the most common causes of oil spill and how they are related to the pipeline type (aboveground, underground or tank). They will also be able to predict average cost of spills according to the pipeline type and the potential cause. This will allow the company to budget a potential spill if it ever happens.

## Data

The dataset that will be used will include a record for each oil pipeline leak or spill reported to the Pipeline and Hazardous Materials Safety Administration since 2010. These records include the incident date and time, operator and pipeline, cause of incident, type of hazardous liquid and quantity lost, injuries and fatalities, and associated costs. This database is available online. [oil spill dataset] (<https://www.kaggle.com/usdot/pipeline-accidents>)

## Outline

* Find most common causes of oil spill
* Find how they relate to the pipeline type
* Find correlation between causes, pipeline type and costs associated

## Deliverables

* R code
* R markdown document