

# Statistical Data Analysis Project

Alexandre De Cuyper

University of A Coruña

February 22, 2024

# Introduction

- ▶ Brief overview of the data analysis project.
- ▶ Importance of the analysis: Understanding the shift in peak positions in crystals of different compositions.
- ▶ Objectives and hypotheses: Investigating if there are significant differences in peak positions among crystals.

# Data Overview

- ▶ Overview of the dataset: Peaks from different crystals with compositions and varying 2 theta values.
- ▶ Key variables and their significance:  
*'2\_theta', Condition, Value, and Cluster.*

# Data Preprocessing

- ▶ Cleaning and handling missing values.
- ▶ Transformation of data for analysis.

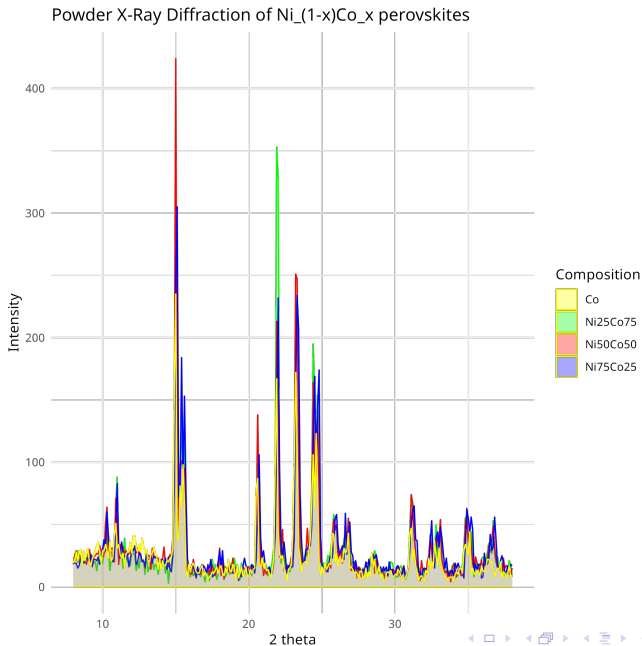
```
> head(data)
  2_theta Ni75Co25 Ni50Co50 Ni25Co75   Co
1    8.00    20.0    18.00    24.00 19.0
2    8.05    19.5    21.98    23.01 20.5
3    8.10    19.0    26.00    22.00 22.0
4    8.15    19.0    27.49    21.50 25.0
5    8.20    19.0    29.00    21.00 28.0
6    8.25    17.0    24.53    24.98 24.5
```

# Exploratory Data Analysis

- ▶ Visualizations and summary statistics.
- ▶ Identification of patterns and trends.

```
1  # Exploratory Data Analysis (EDA)
2  ggplot(data, aes(x = '2_theta')) +
3  geom_ribbon(aes(ymin = 0, ymax = 'Ni25Co75', fill = "Ni25Co75"),
4  alpha = 0.1, color = "green"
5  ) +
6  geom_ribbon(aes(ymin = 0, ymax = 'Ni50Co50', fill = "Ni50Co50"),
7  alpha = 0.1, color = "red"
8  ) +
9  geom_ribbon(aes(ymin = 0, ymax = 'Ni75Co25', fill = "Ni75Co25"),
10 alpha = 0.1, color = "blue"
11 ) +
12 geom_ribbon(aes(ymin = 0, ymax = 'Co', fill = "Co"),
13 alpha = 0.1, color = "yellow"
14 ) +
15 labs(
16   title = "Powder X-ray Diffraction of Ni_(1-x)Co_x perovskites",
17   x = "2_theta",
18   y = "Intensity",
19   fill = "Composition"
20 ) +
21 scale_fill_manual(values = c(
22   "Ni75Co25" = "blue", "Ni50Co50" = "red",
23   "Ni25Co75" = "green", "Co" = "yellow"
24 )) +
25 theme_minimal()
26
```

# Visualization of the Data

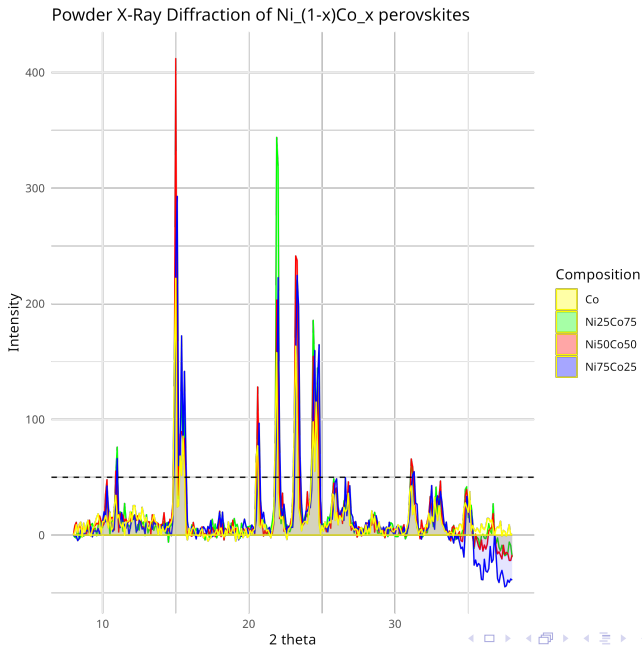


# Exploratory Data Analysis

- ▶ Visualizations and summary statistics.
- ▶ Identification of patterns and trends.

```
1  #Threshold value
2  +
3  geom_hline(
4    yintercept = threshold,
5    linetype = "dashed", color = "black"
6  )
7
```

# Visualization of the Data

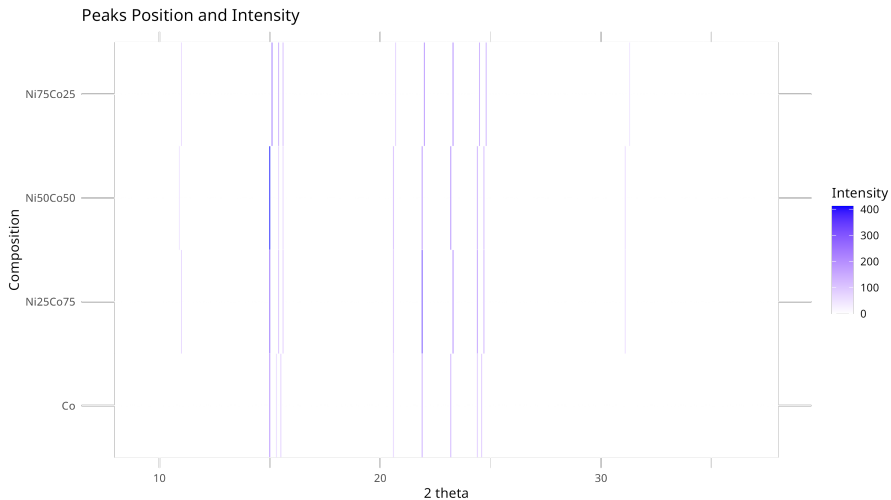




# Peak Identification

```
1 find_peaks <- function(data) {  
2   if (length(data) < 2) {  
3     return(NULL) # No peak in lists with 0 or 1 element  
4   }  
5   peaks <- rep(0, length(data))  
6   # Create a vector to store the value and index of the peaks  
7   for (i in 2:(length(data) - 1)) {  
8     if (data[i] > data[i - 1] && data[i] > data[i + 1]) { # Looking for a  
9       peak  
10      peaks[i] <- data[i]  
11    }  
12  }  
13  # Checking if the first and last value is a peak or not  
14  if (data[1] > data[2]) {  
15    peaks[1] <- data[1]  
16  }  
17  if (tail(data, 1) > tail(data, 2)[1]) {  
18    peaks[length(peaks)] <- data[length(data)]  
19  }  
20  return(peaks)  
}
```

# Peak Identification Results



# Clustering Process

- ▶ Explanation of the clustering algorithm used.
- ▶ Description of the data preparation steps.

```
1 # Your clustering code here
2
```

# Clustering Results

- ▶ Overview of the clustering results.
- ▶ Interpretation of clusters and their characteristics.

cluster\_plot.png

# ANOVA Testing

- ▶ Analysis of Variance (ANOVA) to test hypotheses.
- ▶ Results and interpretation.

```
1 # ANOVA Testing
2 anova_result <- aov('2_theta' ~ Condition *
Cluster, data = data_long)
3 summary(anova_result)
4
```

# Results and Discussion

- ▶ Interpretation of ANOVA results.
- ▶ Comparison of clusters: Assessing the significance of peak position variations.
- ▶ Implications of the findings: Understanding how composition affects peak positions.

# Conclusion

- ▶ Summary of key findings.
- ▶ Limitations and areas for future research.

Any Questions?