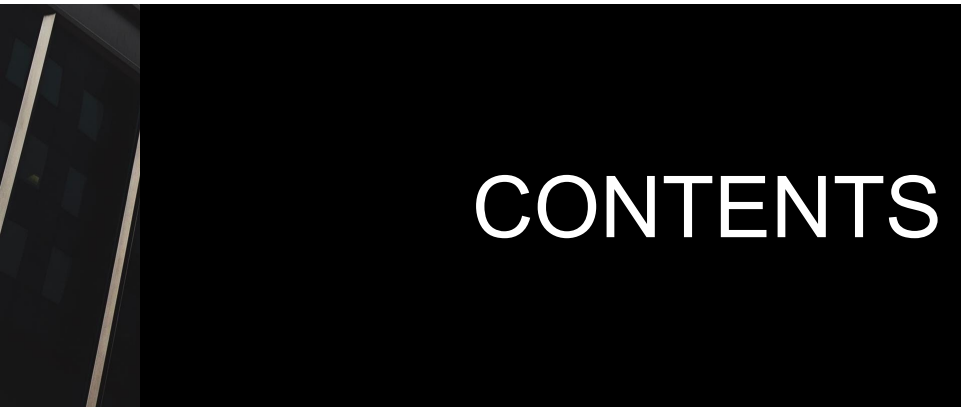
IRIE
Project II
2018

Relation Extraction with The Wall Street Journal

NTU CSIE, Fall 2018 陳信希 教授
顏安孜 陳重吉 {d04922005, d05922016}@ntu.edu.tw



CONTENTS

01 Research Problem

02 Data

03 Task Definition and Evaluation

04 Report

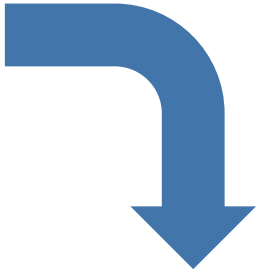
05 Schedule

01 Edge Prediction

Example

Given the following sentence

According to the U.S. Census, almost 10.9 million African Americans, or 28%, **live at or below the poverty line**, compared with 15% of Latinos and approximately 10% of White Americans…….



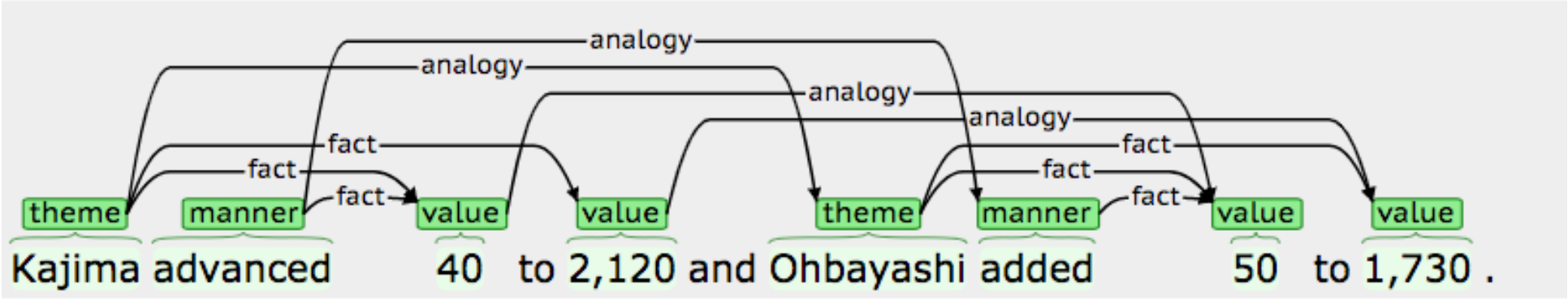
Output

$$\begin{bmatrix} \text{SOURCE} & \text{U.S. Census} \\ \text{QUANT} & \text{live at or below the poverty line} \\ \left\{ \begin{bmatrix} \text{WHOLE} & \text{African Americans} \\ \text{VALUE} & 28\% \end{bmatrix} \right. & \\ \begin{bmatrix} \text{WHOLE} & \text{Latinos} \\ \text{VALUE} & 15\% \end{bmatrix} & \\ \left. \begin{bmatrix} \text{WHOLE} & \text{White Americans} \\ \text{VALUE} & 10\% \end{bmatrix} \right\} & \end{bmatrix}$$

Project Main Task

Edge Prediction

1. Given Tokens (without Nodes information), predict the edge of given pair.
2. Given Tokens and Nodes information, predict the edge of given pair.



More Details

EMNLP 2018

Textual Analogy Parsing: What's Shared and What's Compared among Analogous Facts

.....

FNP 2018

QSRL: A Semantic Role-Labeling Schema for Quantitative Facts

mrlamm/textual-analogy-parsing

https://github.com/mrlamm/textual-analogy-parsing

.....

02 Data

Data - Overview

Train – 1000 Sentences

Test – 97 Sentences

Three part for each sentence

- tokens
- nodes
- edges

Data - Annotation

Tokens

Use the tokens order to match index in Nodes and Edges

Nodes

Total 28470 nodes are annotated with following 18 kinds of labels.

value, agent, condition, theme, theme_mod, quant_mod, co_quant, null, location, whole, source, reference_time, quant, manner, time, cause, +, -

Edges

They annotated 18620 edges with 3 kinds of labels:

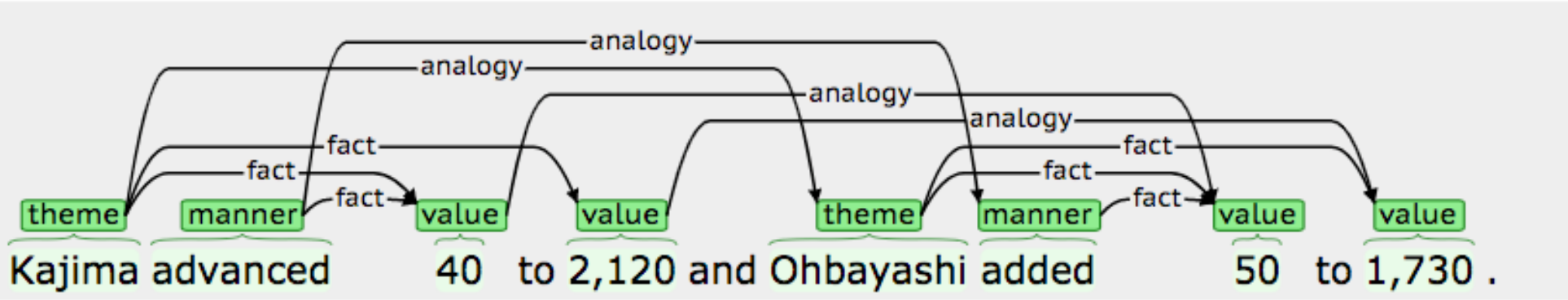
equivalence, fact, analogy

Data - Example

‘tokens’ : [‘Kajima’ , ‘advanced’ , ‘40’ , ‘to’ , ‘2,120’ , ‘and’ , ‘Ohbayashi’ , ‘added’ , ‘50’ , ‘to’ , ‘1,730’ , '.']

'nodes': [[[0, 1], {'theme': 1.0}, {'null': 1.0}, {'null': 1.0}],
[[1, 2], {'manner': 1.0}, {'null': 1.0}, {'null': 1.0}],
[[2, 3], {'value': 1.0}, {'null': 1.0}, {'+': 1.0}],
[[4, 5], {'value': 1.0}, {'null': 1.0}, {'null': 1.0}],
[[6, 7], {'theme': 1.0}, {'null': 1.0}, {'null': 1.0}],
[[7, 8], {'manner': 1.0}, {'null': 1.0}, {'null': 1.0}],
[[8, 9], {'value': 1.0}, {'null': 1.0}, {'+': 1.0}],
[[10, 11], {'value': 1.0}, {'null': 1.0}, {'null': 1.0}]]

'edges': [[[0, 1], [2, 3], {'fact': 1.0}],
[[0, 1], [4, 5], {'fact': 1.0}],
[[0, 1], [6, 7], {'analogy': 1.0}],
[[1, 2], [2, 3], {'fact': 1.0}],
[[1, 2], [7, 8], {'equivalence': 1.0}],
[[2, 3], [8, 9], {'analogy': 1.0}],
[[4, 5], [10, 11], {'analogy': 1.0}],
[[6, 7], [8, 9], {'fact': 1.0}],
[[6, 7], [10, 11], {'fact': 1.0}],
[[7, 8], [8, 9], {'fact': 1.0}]]}



03 Task Definition and Evaluation

Task & Evaluation

Edge Prediction

1. Given Tokens (without Nodes information), predict the edge of given pair.
2. Given Tokens and Nodes information, predict the edge of given pair.

Please report Precision, Recall, and F1

- We use F1 to evaluate the performance

04 Report

Report and Code

Language

- Chinese or English (Be clear in meaning!).

Pages

- No more than 6 pages (with readable font size)

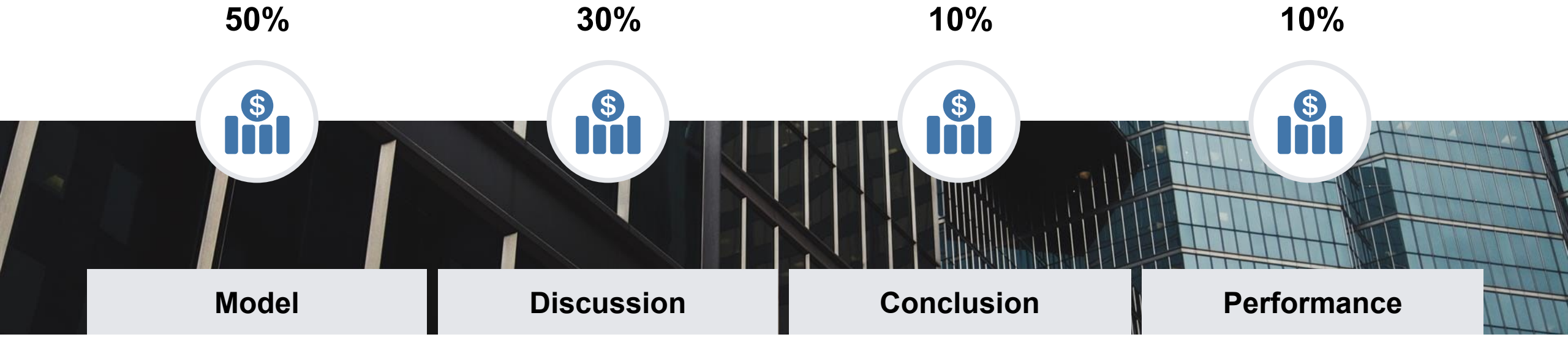
Format

- Zip (upload to Ceiba)
  - Report: PDF
  - Code: Github link or folder
    - Please provide ReadMe file

Must include

- Name and ID
- Division of work
- Methods
- Evaluation
- Discussion
- Conclusion

Grade



50%	30%	10%	10%
Model	Discussion	Conclusion	Performance

Try Three models

Schedule

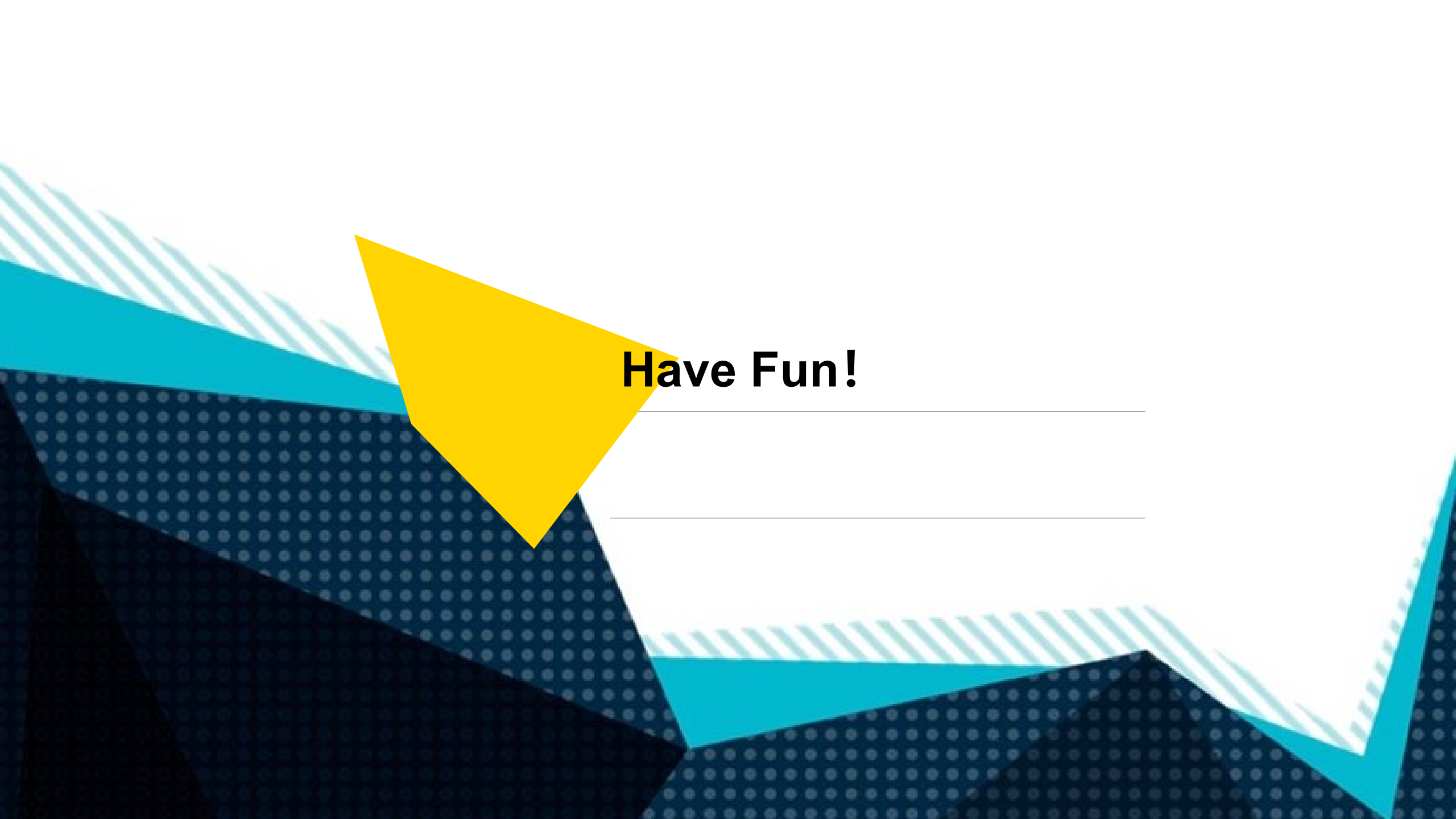
12/13
Project Slides Release

1/10

Final Exam

1/11 23:59

Report & Code Submission Due


Have Fun!