POLYTECHNIQUE MONTRÉAL

LE GÉNIE EN PREMIÈRE CLASSE

# On the Influence of Representation Type and Gender on Program Comprehension

Zohreh Sharafi

Département de Génie Informatique et Génie Logiciel

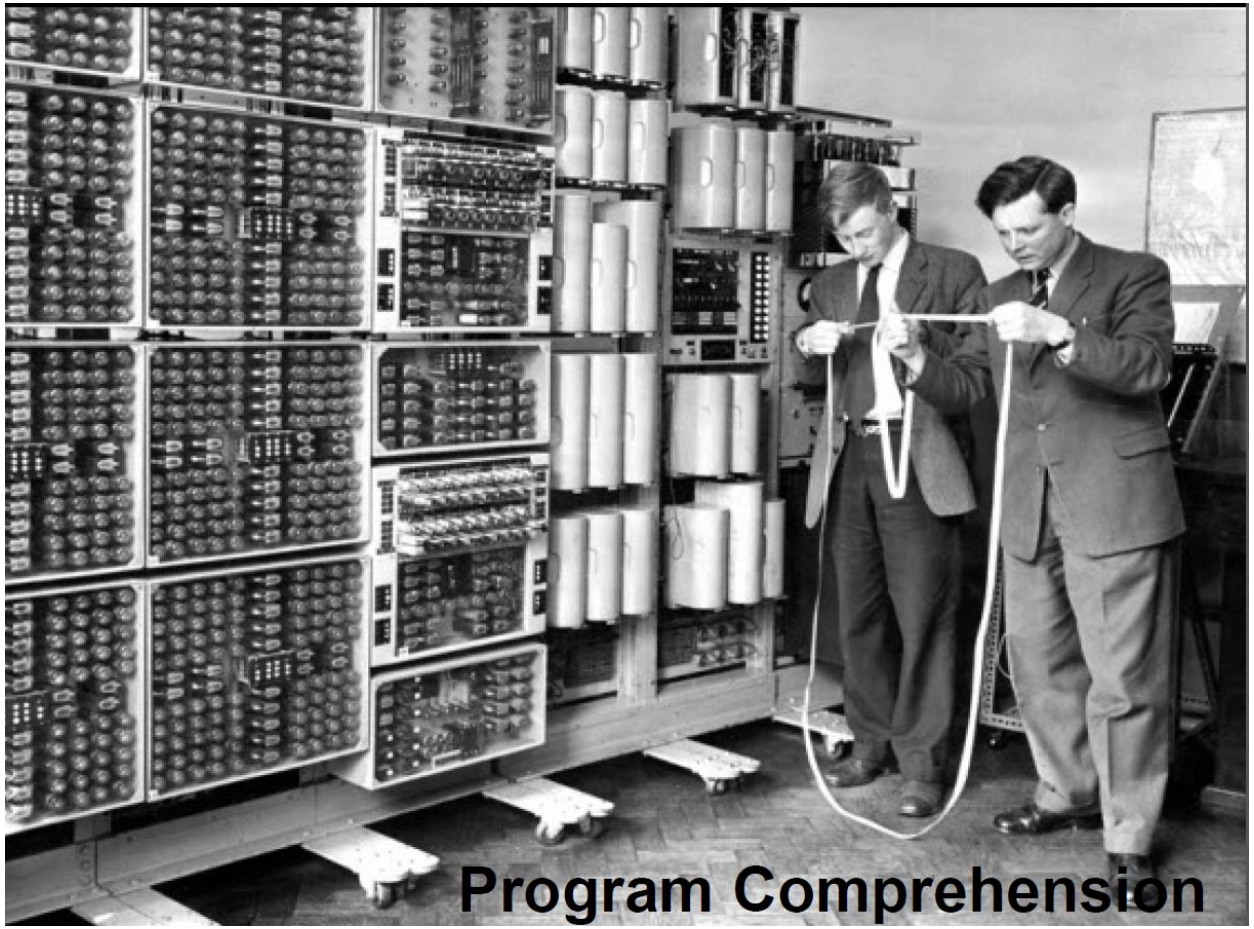
Polytechnique Montréal

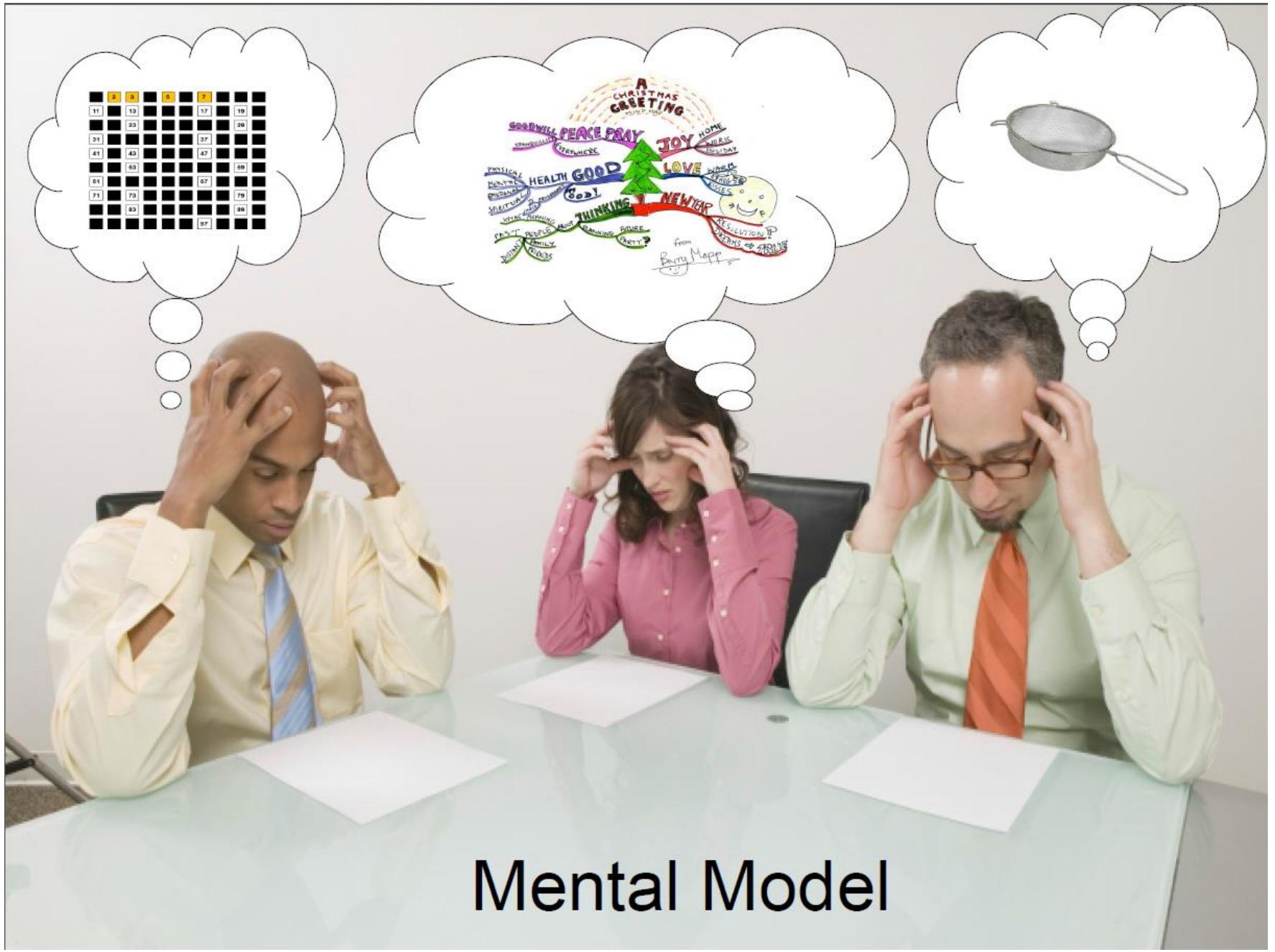
Soccerlab
SOftware Cost-effective Change and Evolution Research Lab

Ptidej

# Program Comprehension

"Programming is far more complex than usual human mental activities studied by psychologists" (Weinberg and Schulman, 1974)




A CHRISTMAS GREETING
GOODWILL
PEACE
PRAY
JOY
HOME
HEALTH
GOOD
LOVE
GOD
THINKING
NEW YEAR
RESOLUTION
from Barry Mapp
Mental Model

# Program Comprehension

- Over 35 years of research
- Rich with diverse theories


Think-aloud

Observational studies


Questionnaires

# Eye-tracking Technology

**EyeLink II**

(SR Research)

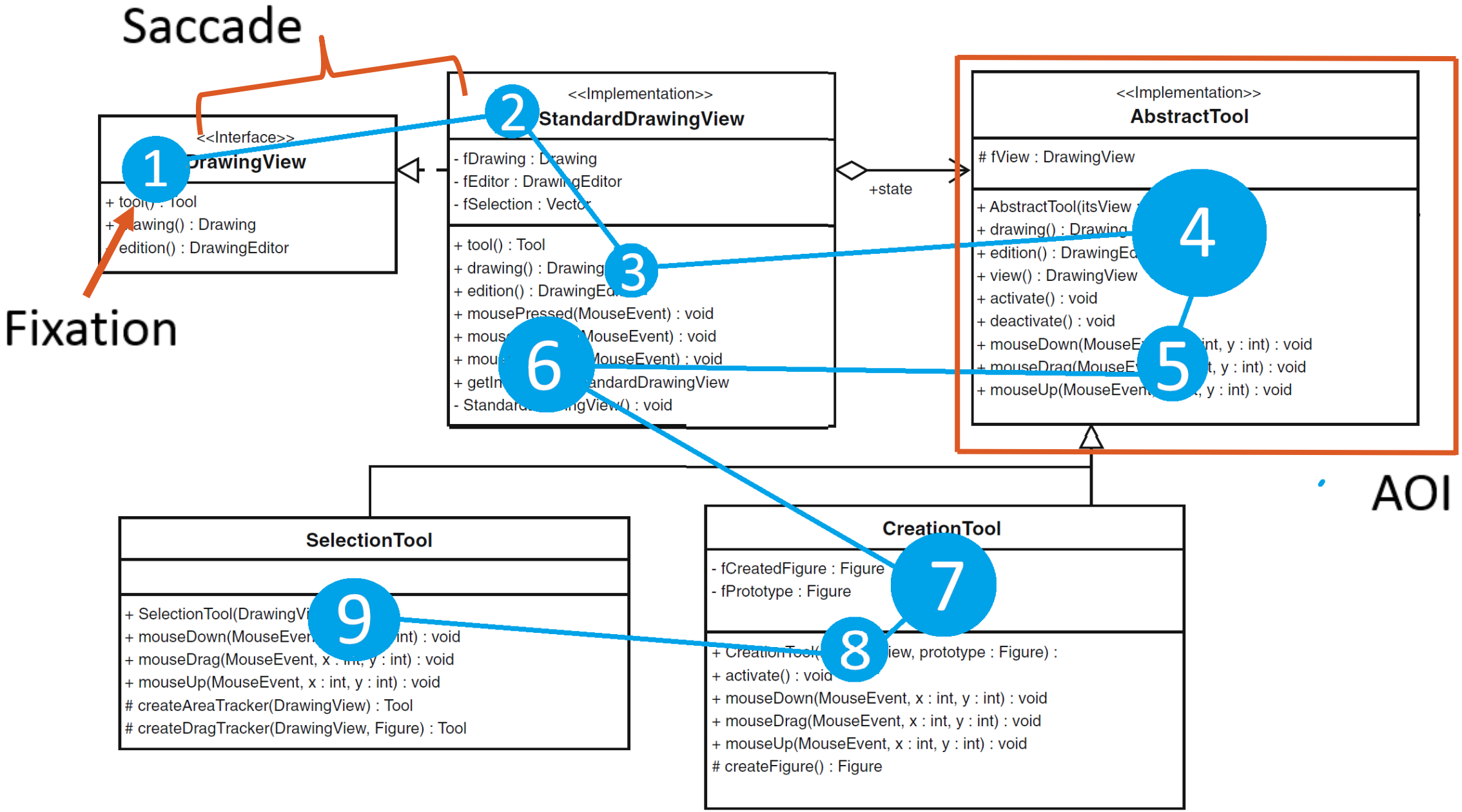
**faceLAB 5**

(Seeing Machines)

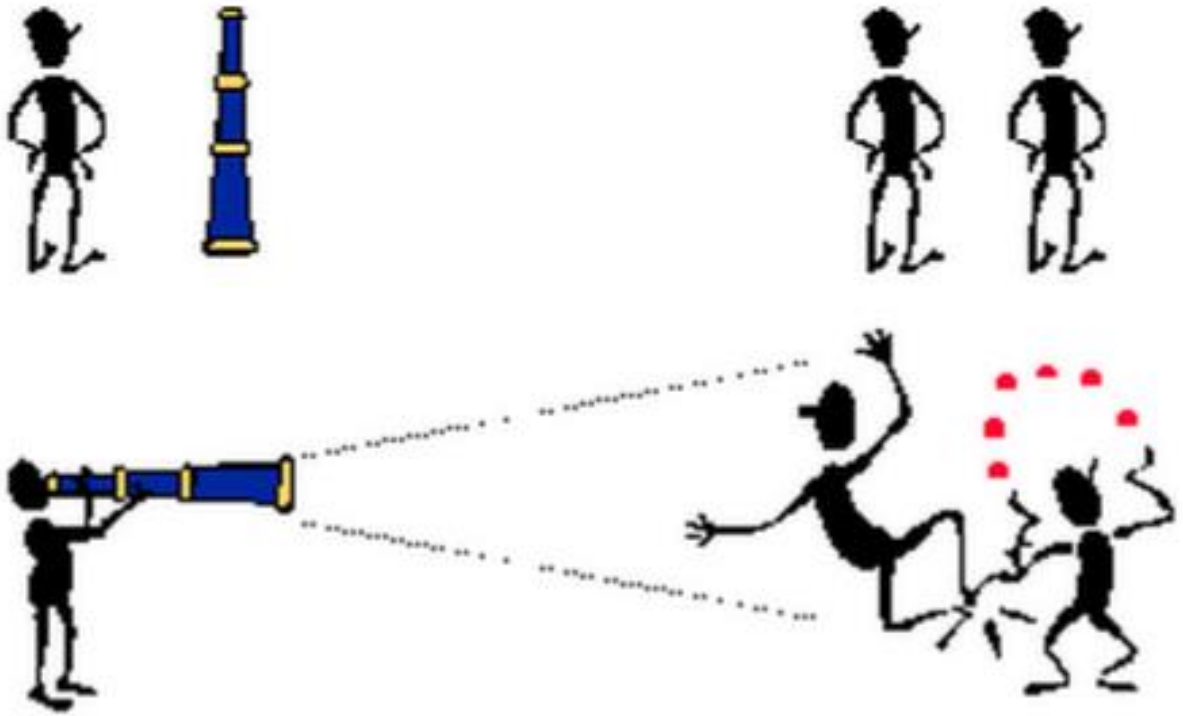
**SMI Eye Tracking Glasses**

(Sensomotoric Instruments)

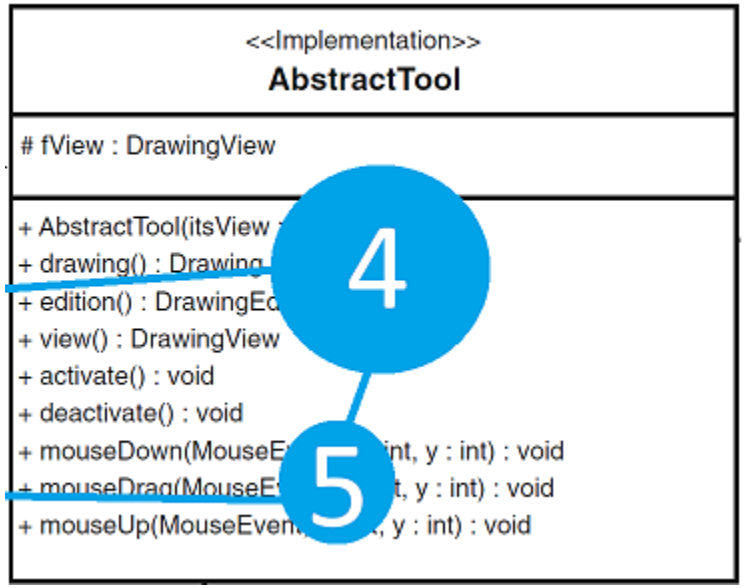

Which objects are drawn by one instance of the class?
SAMSUNG




Scan-path = 1,2,3,4,5,6,7,8,9



Hawthorn Effect



Representation's type

Developer's characteristics


Program's characteristics

Task variability

**Program Comprehension**

# Representation Type (Graphical vs. Textual )

Vs.

Apple

---

- Round
- Red or Yellow
- Edible fruit

# Developers' Characteristics

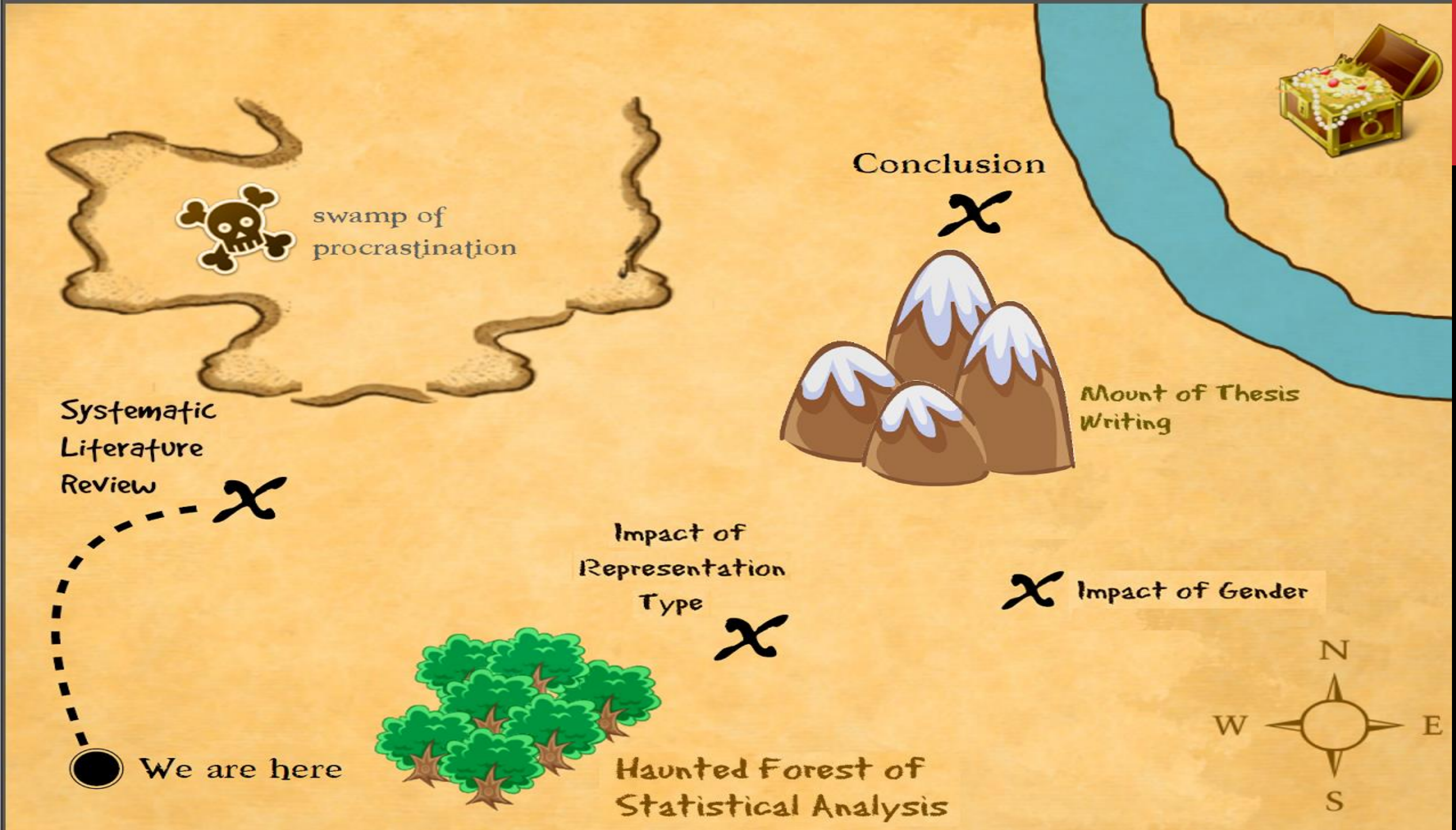
- Expertise (novices vs. experts)

There is still a great divergence between developers' abilities that cannot be detected only by their level of expertise. (Storey, 2005).

# Thesis Statement

Representation type and gender:

- To impact the cognitive process
- To impact developers' efficiency and effectiveness
- To be proxy for developers' viewing strategies
    - To be inferred partly from the developers' eye-movements


swamp of procrastination
Conclusion
Mount of Thesis Writing
Systematic Literature Review
We are here
Impact of Representation Type
Impact of Gender
Haunted Forest of Statistical Analysis
N
W
E
S

# Systematic Literature Review (SLR)

Evidence-based Paradigm

1992

Evidence-based Software Engineering (EBSE)

2004


SLR about SLRs in SE

2010

UML
Unit Testing
Theory

A Systematic Literature Review on the Usage of Eye-tracking in Software Engineering

# A Systematic Literature Review
## Descriptive Statistics

**How many**

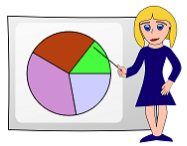
- ✓ 1990 - 2014
- ✓ 36 articles
  - ✓ 9 journal
  - ✓ 2 workshop
  - ✓ 25 conference

Under revision, IST Journal, 2013

# A Systematic Literature Review
## Topics

- Code comprehension (12)
- Model comprehensions (10)
- Debugging (9)
- Collaborative interaction (3)
- Traceability (2)





# A Systematic Literature Review
## Metrics

- Effort:
  - Fixation, Saccades
  - Visited AOIs
- Visual gaze behaviour
- Scan-path

# A Systematic Literature Review
## Limitation

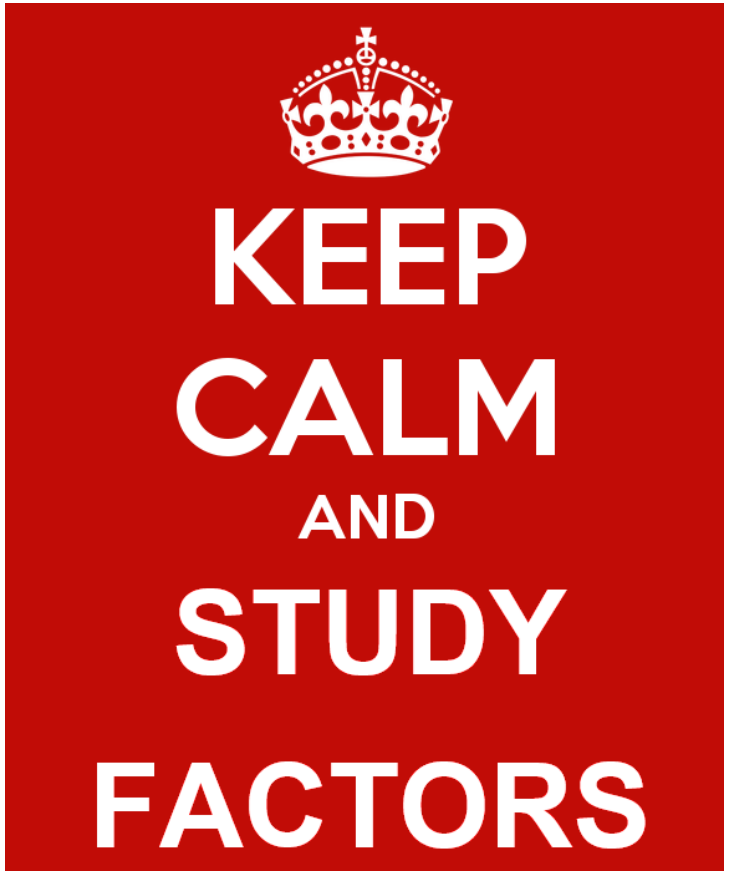
- Technology
- Data analysis
- Participant & material selection
- Experimental setting

# A Systematic Literature Review
## Conclusion

- To provide descriptive statistics
- To present an annotated bibliography
- To summarise all the metrics and tools
- To present limitations
- To provide a unified terminology for reporting

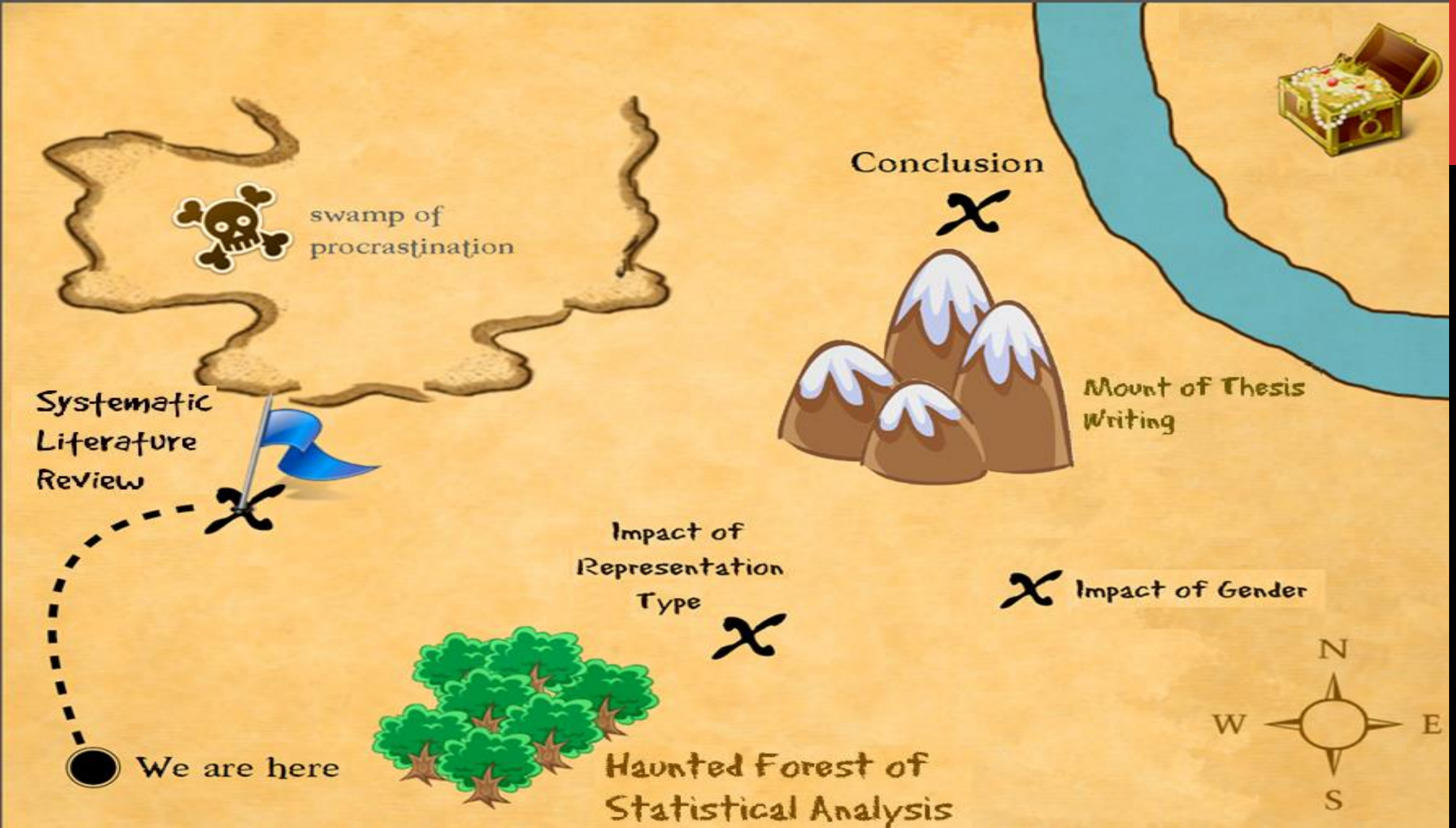
# A Systematic Literature Review

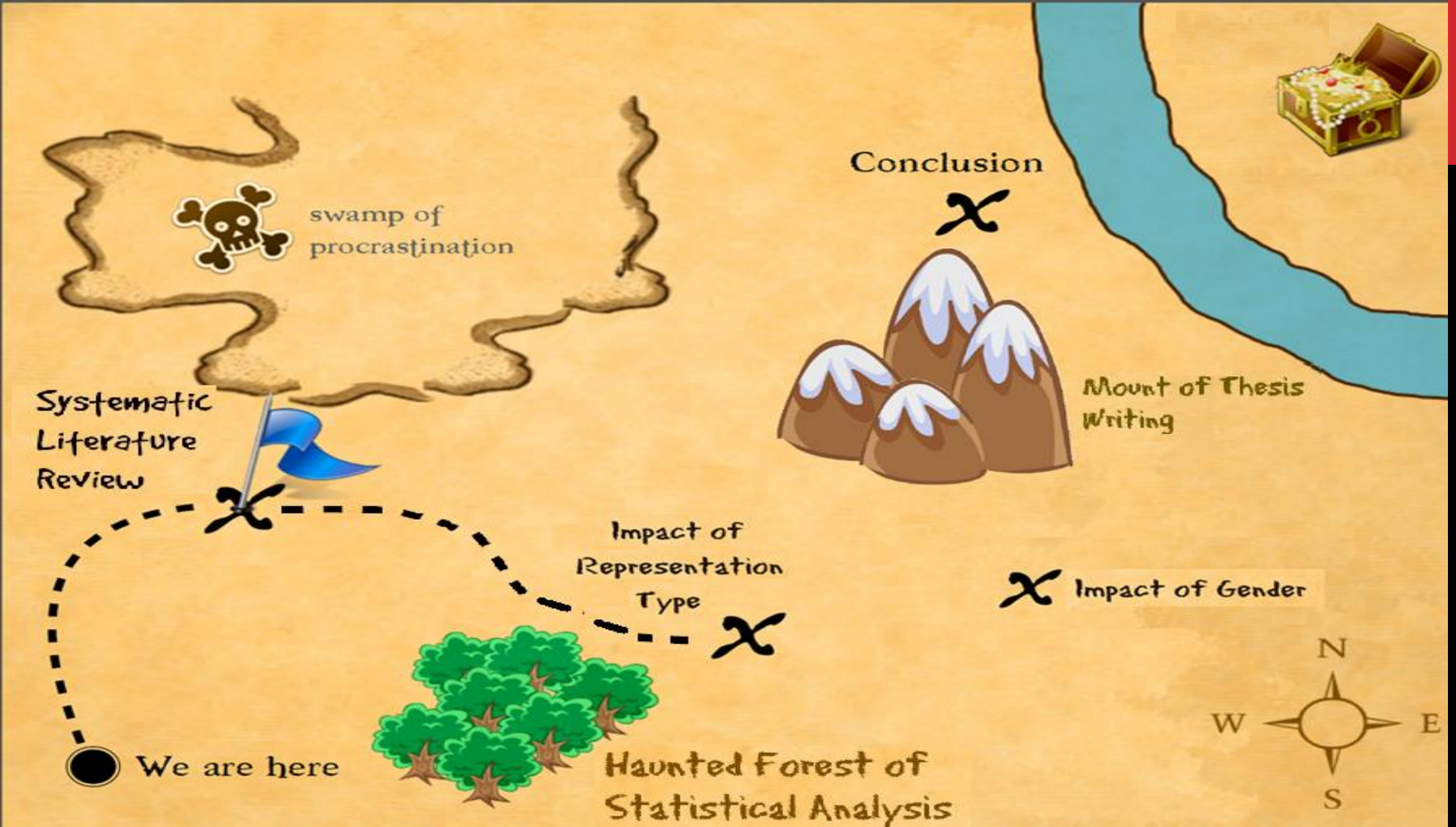
## Lesson Learned



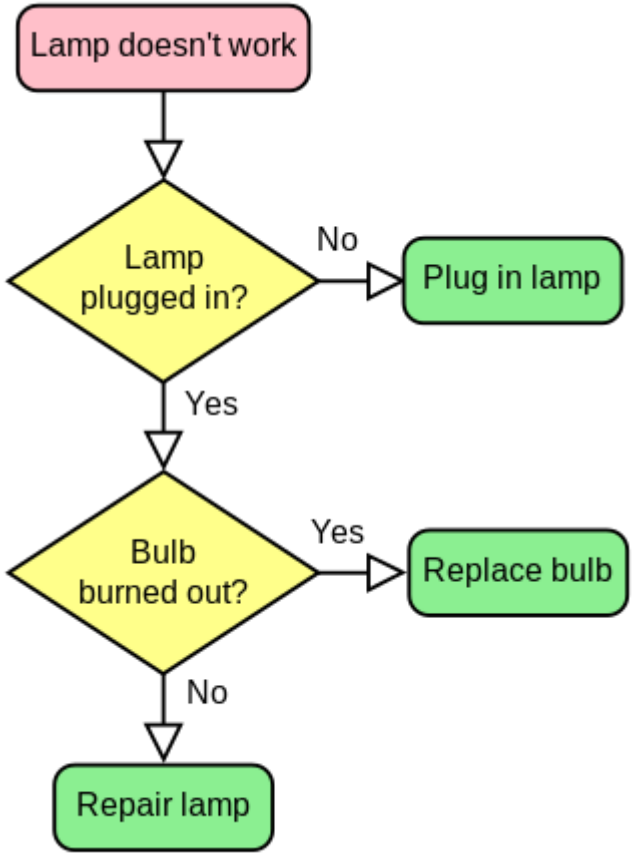

**USE METRICS**

**MITIGATE LIMITATIONS**


swamp of procrastination
Conclusion
Mount of Thesis Writing
Systematic Literature Review
We are here
Impact of Representation Type
Impact of Gender
Haunted Forest of Statistical Analysis
N
W
E
S


swamp of procrastination
Conclusion
Mount of Thesis Writing
Systematic Literature Review
Impact of Representation Type
Impact of Gender
We are here
Haunted Forest of Statistical Analysis
N
W
E
S

# Picture Is Worth a Thousand Word




UNIFIED MODELING LANGUAGE

# The Impact of Representation Type

Only a few works compare textual vs. graphical representations for program comprehension

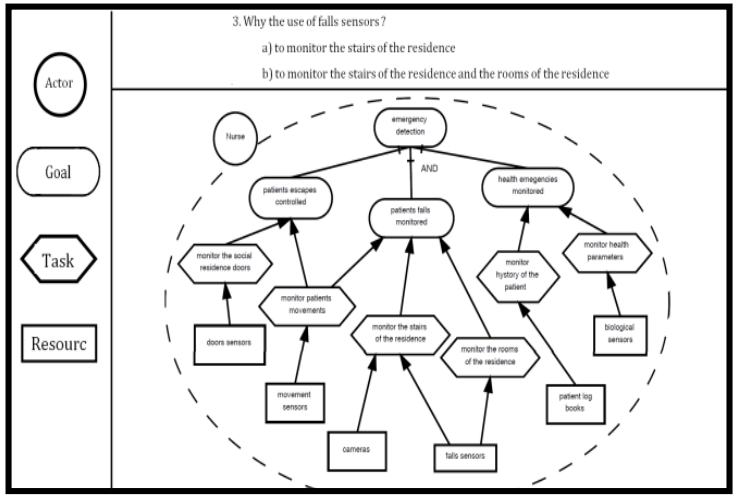
RQ1: Developers' preferred representation

RQ2: Impact of representation type of time, accuracy, and effort

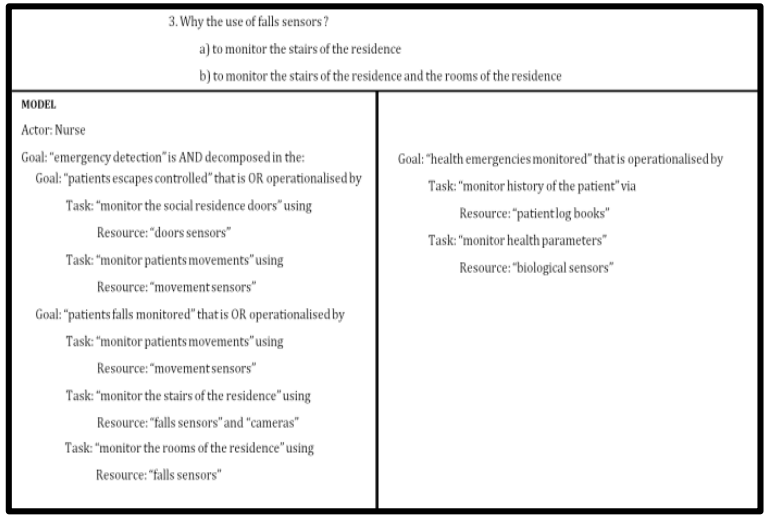
# The Impact of Representation Type

## Task

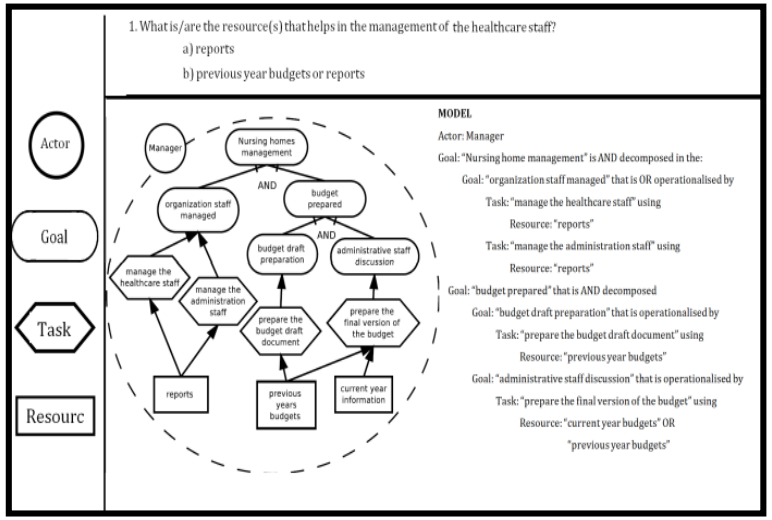
### Session 1

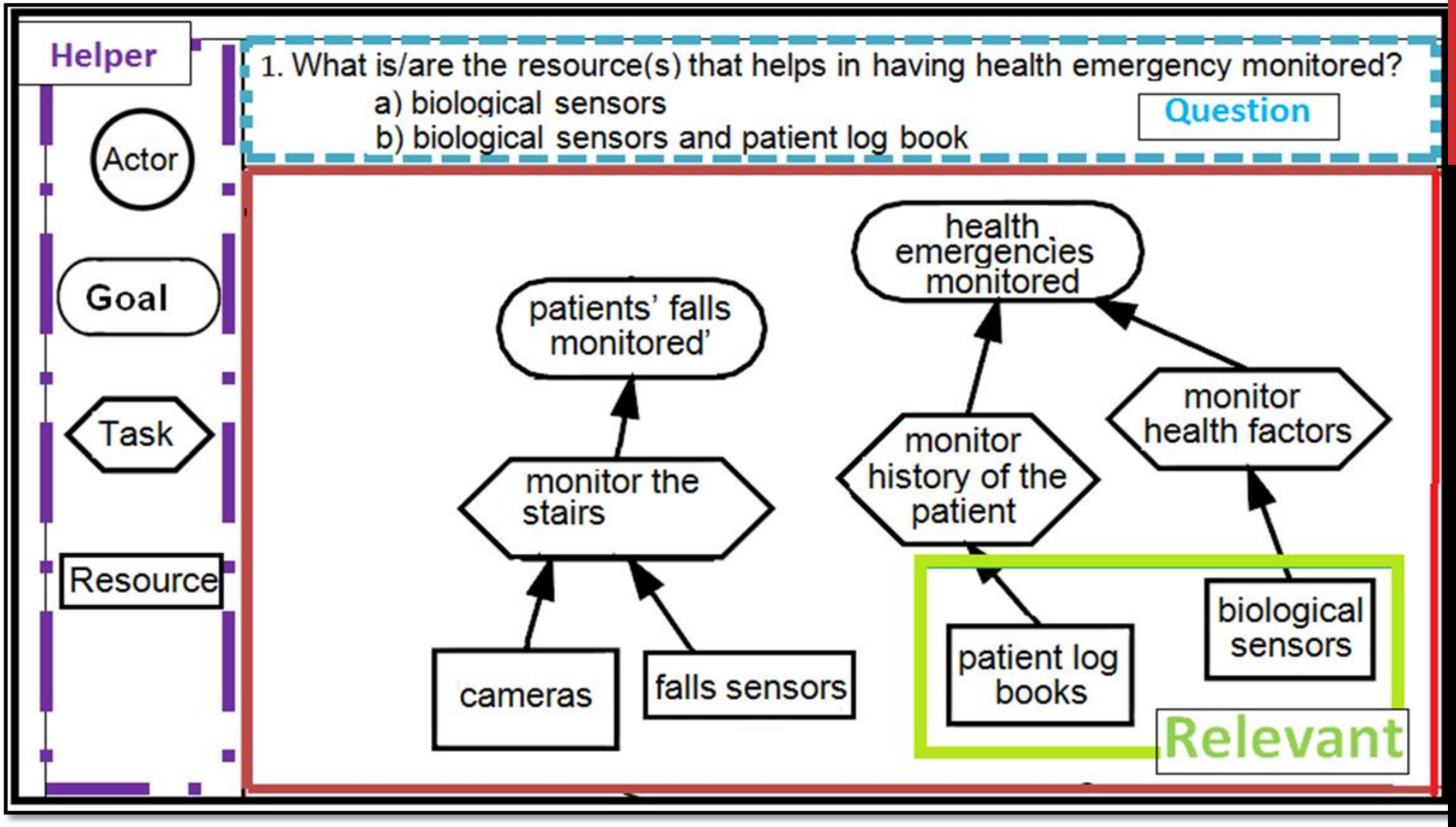



### Session 2




### Session 3




Helper
Actor
Goal
Task
Resource
1. What is/are the resource(s) that helps in having health emergency monitored?
a) biological sensors
b) biological sensors and patient log book
Question
health emergencies monitored
patients' falls monitored'
monitor the stairs
monitor history of the patient
monitor health factors
cameras
falls sensors
patient log books
biological sensors
Relevant

1. What is/are the resource(s) that helps in having health emergency monitored?
   a) biological sensors
   b) biological sensors and patient log book

Question

Goal: "patients' falls monitored" that is OR operationalised by

Task: "monitor the stairs" using

Resource: "falls sensors" and "cameras"

Model

Goal: "health emergencies monitored" that is operationalised by

Task: "monitor history of the patient" via

Resource: "patient log books"

Task "monitor health factors"

Resource: "biological sensors"

Relevant

# The Impact of Representation Type

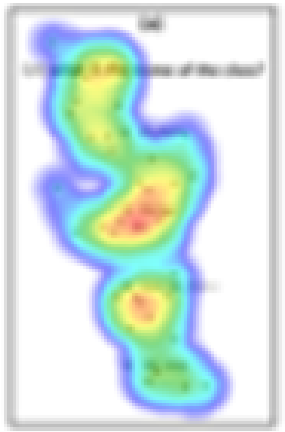
## Participants

Total: 28

| Academic Background | | | Gender | |
|---|---|---|---|---|
| **PhD** | **Masters** | **BSc** | **Male** | **Female** |
| 15 | 11 | 2 | 16 | 12 |

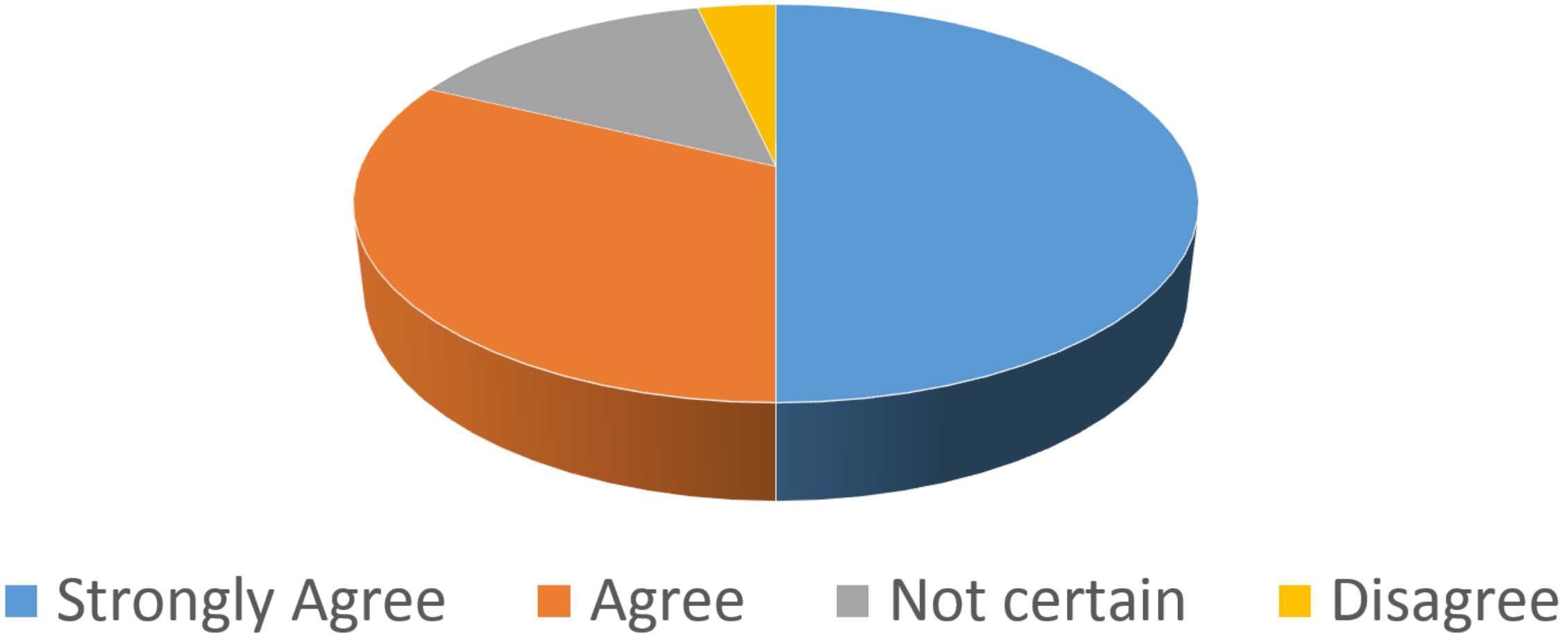
faceLAB

TAUPE



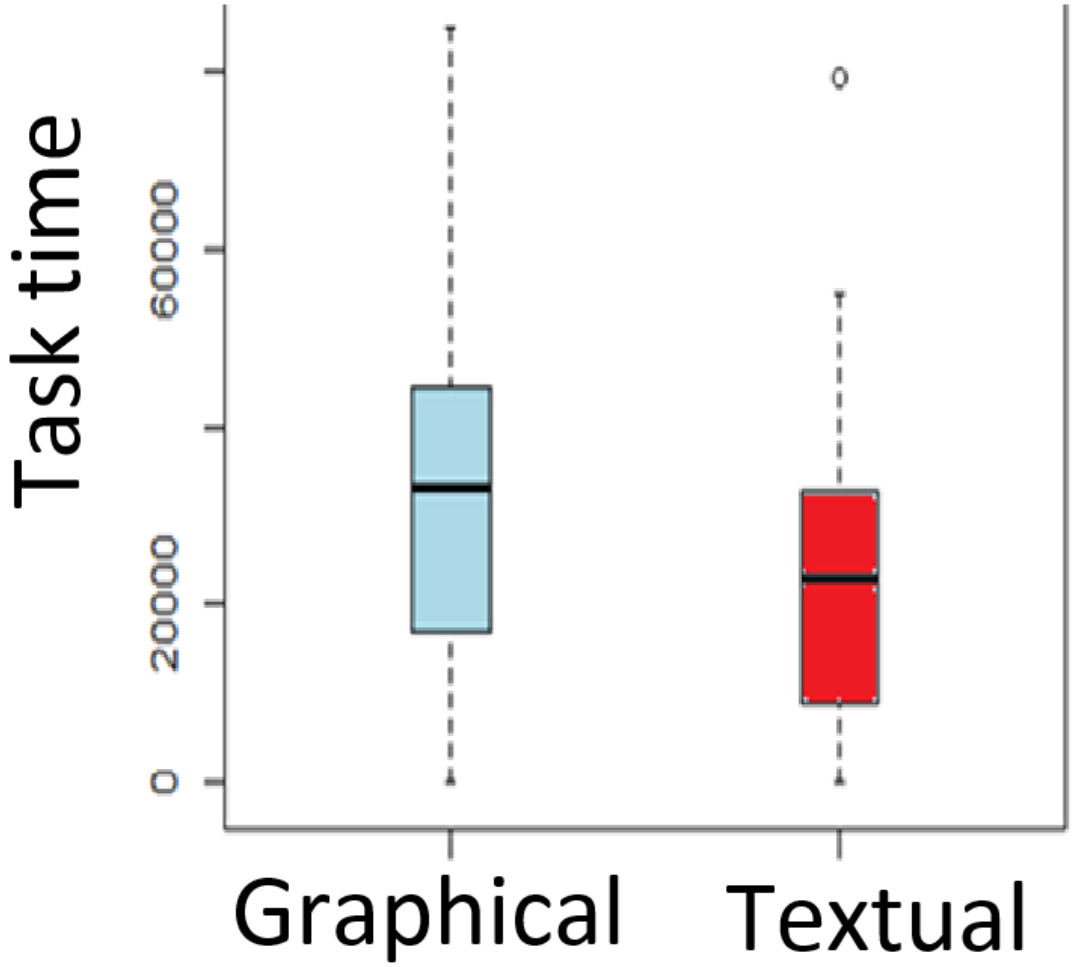
MATLAB

# RQ1: Is Graphical  More Effective?

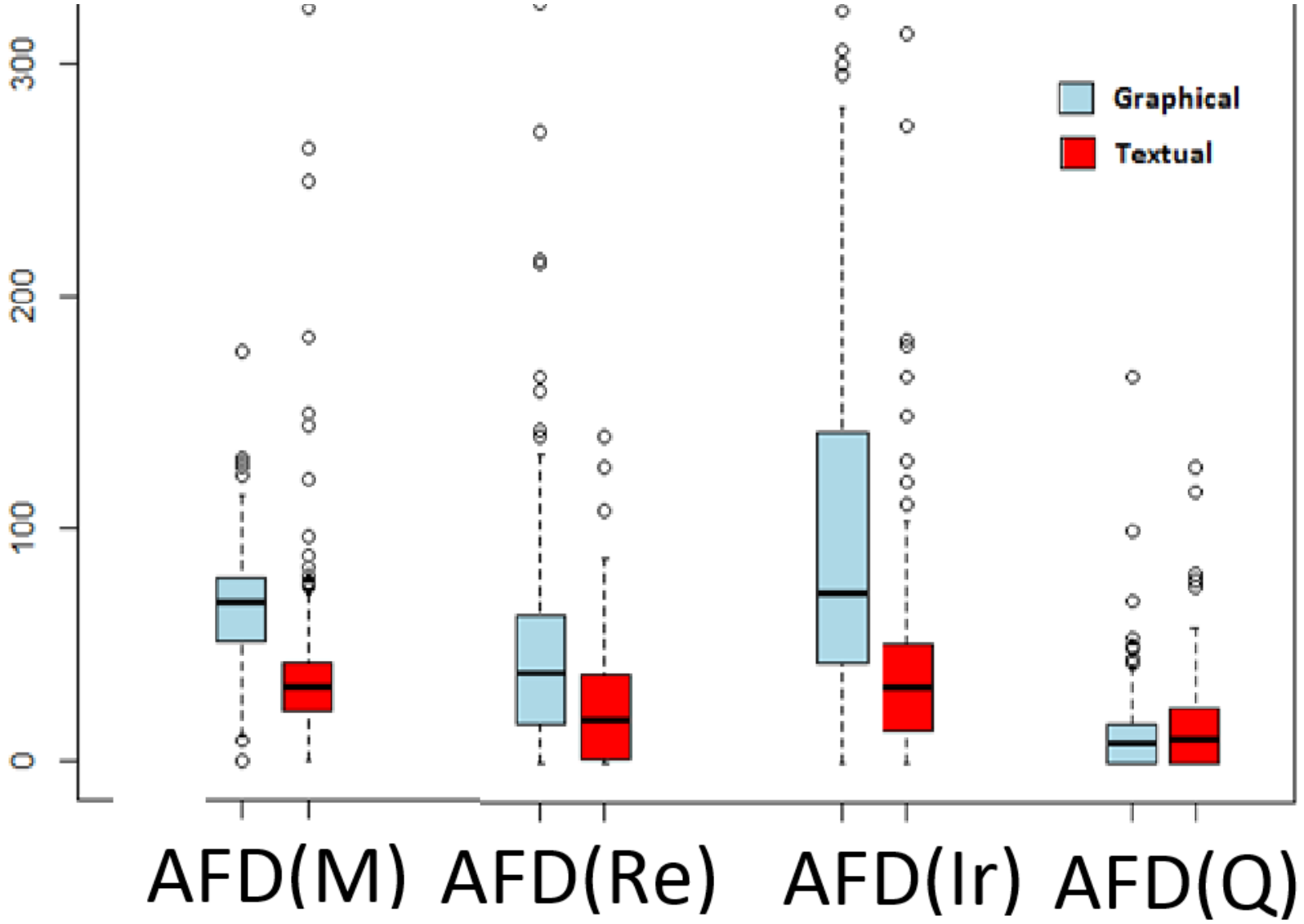


# RQ2: Impacts on Developers' Efficiency

| | Accuracy % | |
|---|---|---|
| | **Correct** | **Wrong** |
| **Graphical** | 97% | 3% |
| **Textual** | 98% | 2% |
| **Mixed** | 96% | 4% |

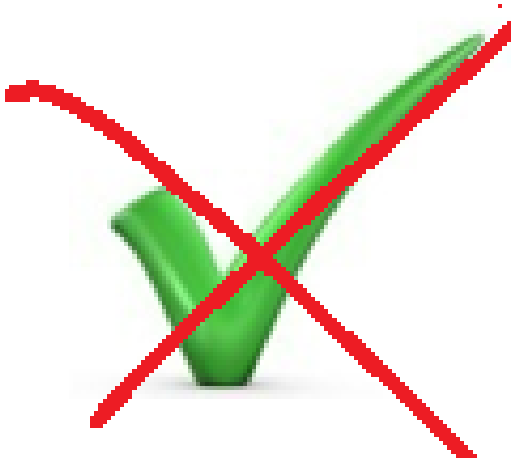


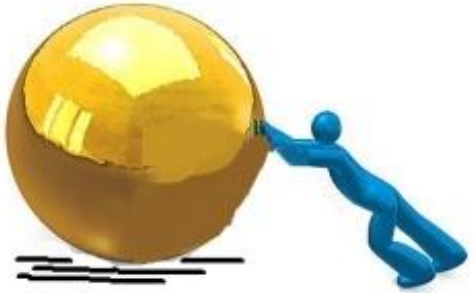
# RQ2: Impacts on Developers' Efficiency




M: Model
Re: Relevant
Ir: Irrelevant
Q: Question

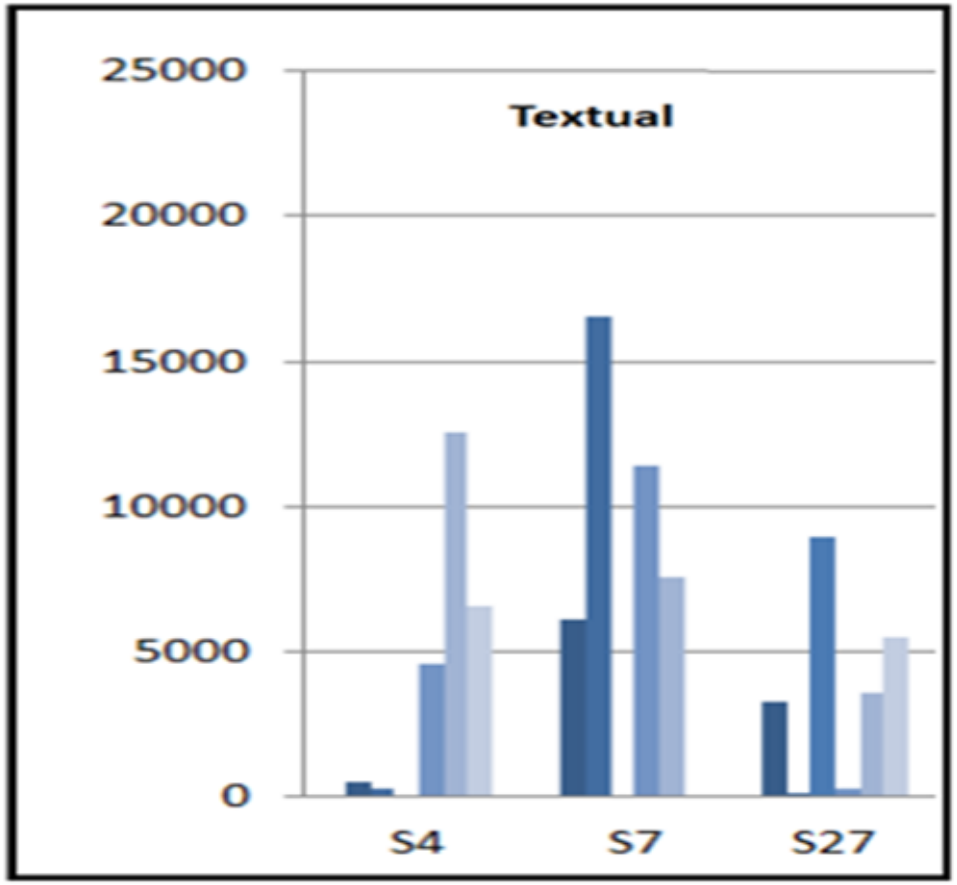
# The Impact of Representation Type
## Conclusion

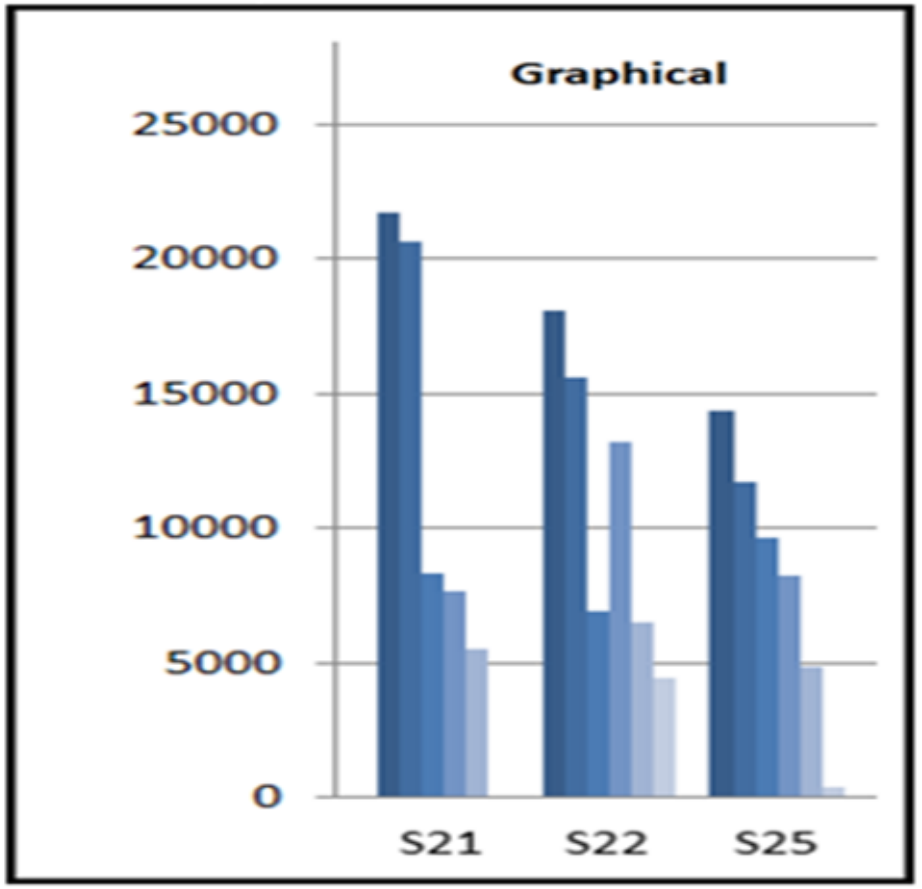




# The Impact of Representation Type
## Conclusion

- Prefer graphical
- Work with graphical in mixed stimulus
- Find relevant parts faster

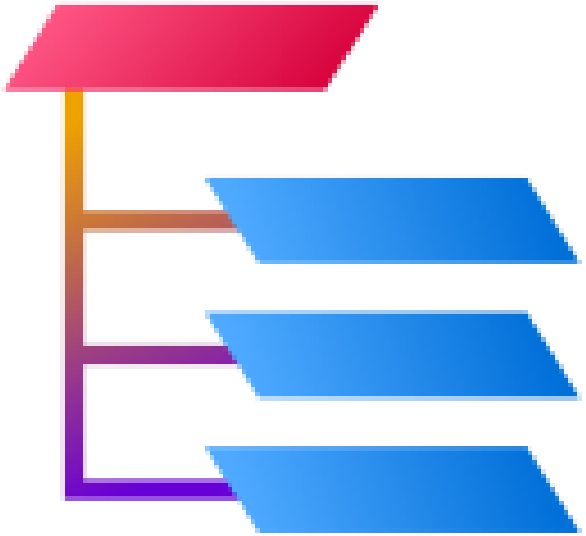



Start time of the first fixation on relevant AOIs

# The Impact of Representation Type

## Lesson Learned

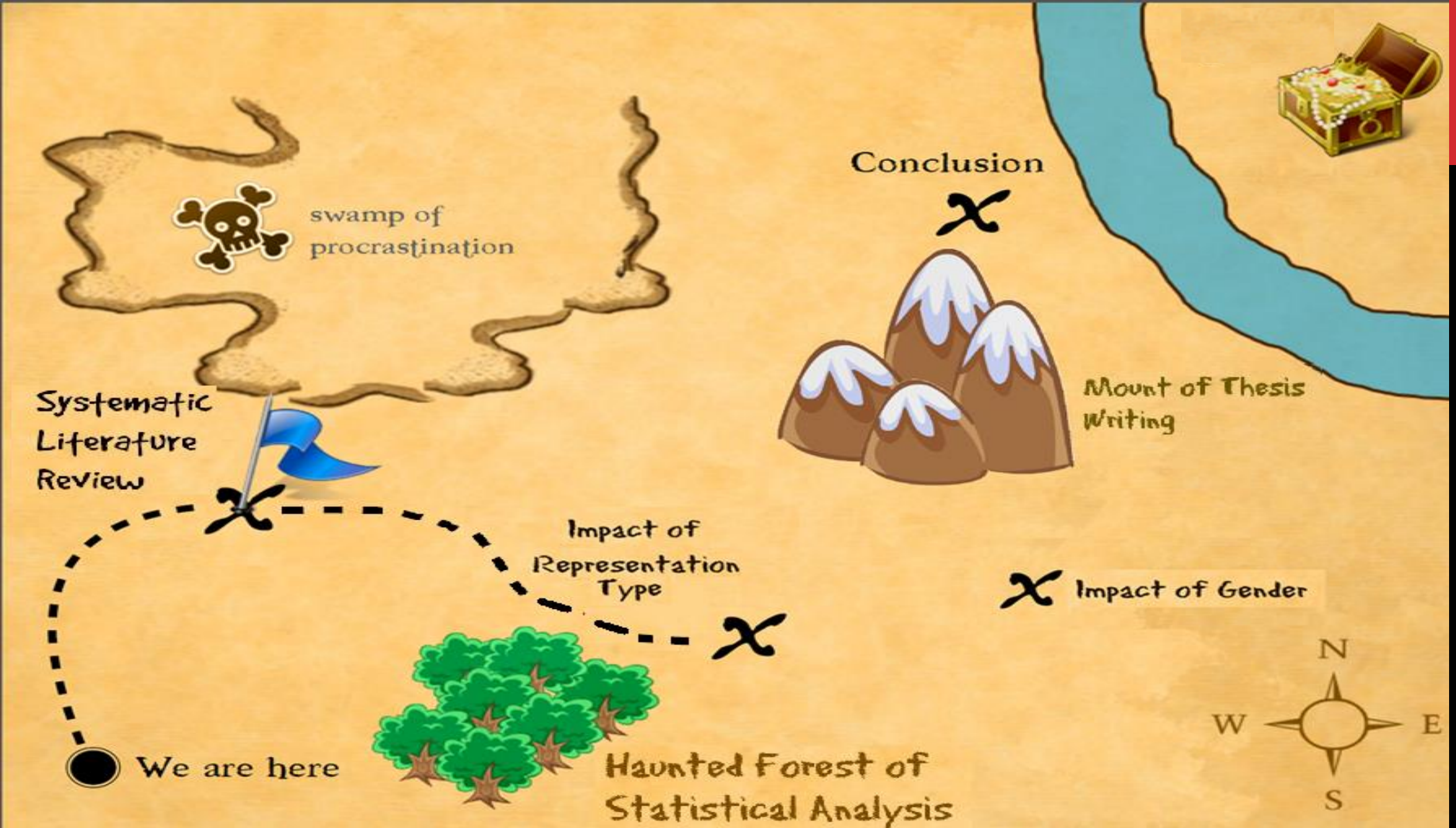


Importance of Layout

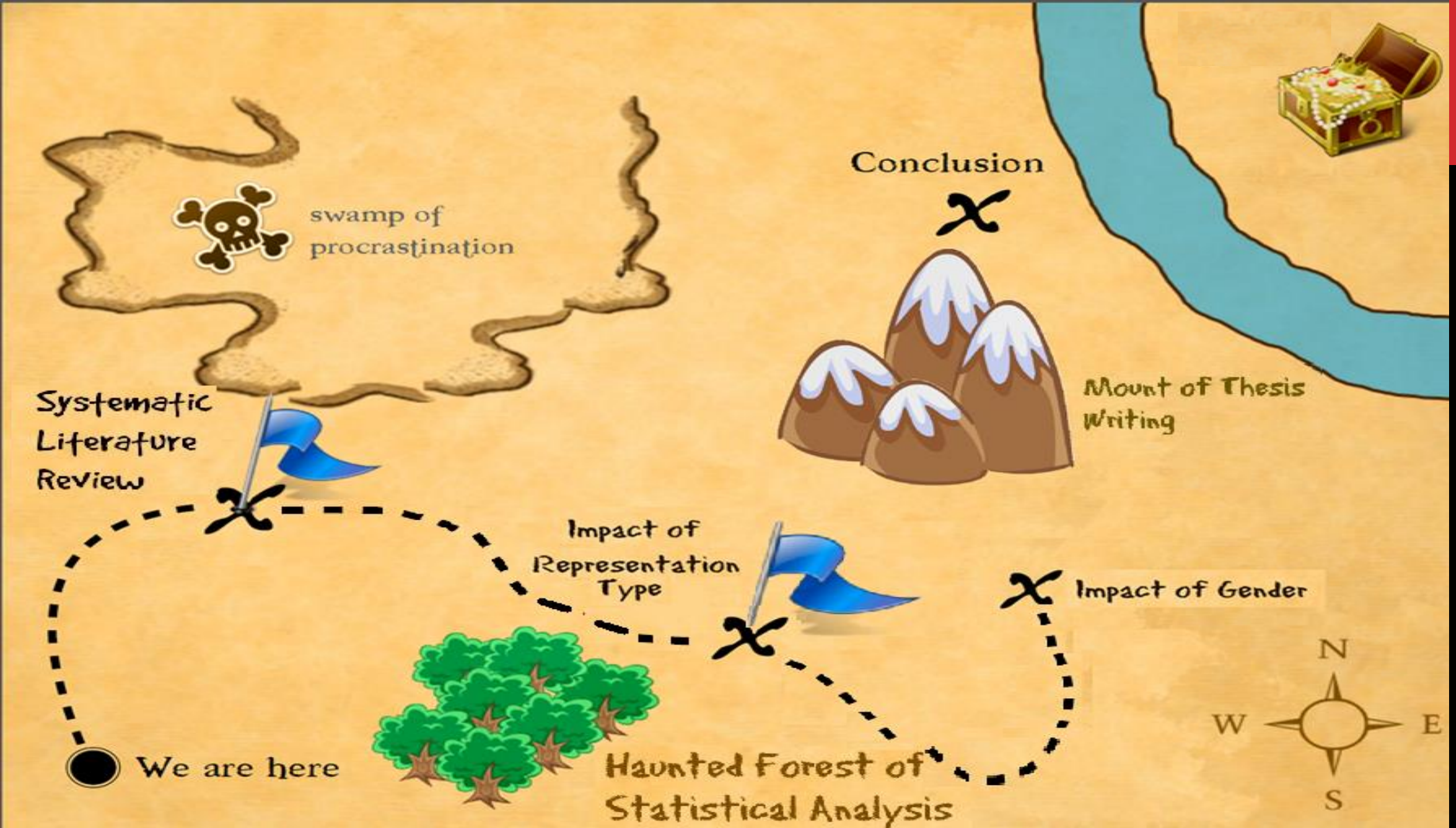
```
/**
 * Simple HelloButton() method.
 * @version 1.0
 * @author john doe <doe.j@example.com>
 */
HelloButton()
{
  JButton hello = new JButton( "Hello, wor
  hello.addActionListener( new HelloBtnList

  // use the JFrame type until support for t
  // new component is finished
  JFrame frame = new JFrame( "Hello Button"
  Container pane = frame.getContentPane();
  pane.add( hello );
  frame.pack();
  frame.show();           // display the fra
}
```

Source code as structures text


swamp of procrastination
Conclusion
Mount of Thesis Writing
Systematic Literature Review
Impact of Representation Type
Impact of Gender
We are here
Haunted Forest of Statistical Analysis
N
W
E
S


swamp of procrastination
Conclusion
Mount of Thesis Writing
Systematic Literature Review
Impact of Representation Type
Impact of Gender
We are here
Haunted Forest of Statistical Analysis
N
W
E
S


swamp of procrastination
Conclusion
Mount of Thesis Writing
Systematic Literature Review
Impact of Representation Type
Impact of Gender
We are here
Haunted Forest of Statistical Analysis
N
W
E
S

# Women Take a Wider View

M. Czerwinski, D. S. Tan, and G. G. Robertson, “Women take a wider view,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ’02. New York, NY, USA: ACM, 2002,

# The Impact of Gender
## Why Do We Care?

- Studying the need of one sub-population can benefit both sides
- Design tools better adapted to different developers
- Support different program understanding strategies

# The Impact of Gender

## Experiment 1: The Impact of Identifier Style

✓Previous work, such as (Binkley et al., 2012) and (Sharif and Maletic, 2010c), reported contradictory findings.

RQ1: The impact of identifier style on effort, the task time, and accuracy

RQ2: The impact of gender on effort, the task time, accuracy, and viewing strategies

# The Impact of Gender
## Experiment 1: Task

Source code stimulus

```
public class PrimeNumCalc {

    private final int UpperLimit = 100;
    public void calcPrimeNums() {
        int i = 0;
        int nCounter = 0;

        while (++i <= UpperLimit) {
            int innerLoop = (int) Math.ceil(Math.sqrt(i));
            boolean isPrimeNumber = false;

            while (innerLoop > 1) {
                if ((i != innerLoop) && (i % innerLoop == 0)) {
                    isPrimeNumber = false;
                    break;
                } else if (!isPrimeNumber) {
                    isPrimeNumber = true;
                }
                --innerLoop;
            }
            if (isPrimeNumber) {
                System.out.println(i);
                ++nCounter;
            }
        }
        System.out.println("Nr of prime numbers found: " + NCounter);
    }
}
```

**What is the output of this program?**

Question stimulus

**Q1: what is the name of the class?**

1. **PrimeNumClac**
2. **NumCalculator**
3. **myClass**
4. **PrClass**

# The Impact of Representation Type

## Participants

Total: 24

| Academic Background | | | Gender | |
|---|---|---|---|---|
| **PhD** | **Masters** | **BSc** | **Male** | **Female** |
| 11 | 10 | 3 | 15 | 9 |

# The Impact of Gender

## Experiment 1: Results

| | Accuracy | Time (min) |
|---|---|---|
| **Male** | 74% | 5.94 |
| **Female** | 82% | 7.18 |

(a)

Q4) what is the name of the class?

1) 2D_frame

2) Painter

3) Java_2D_frame

4) My_class

(b)

Q4) what is the name of the class?

1) 2D_Frame

2) Painter

3) Java_2D_Frame

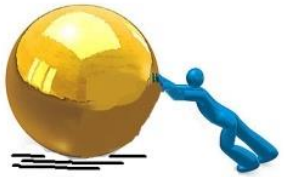
4) My_Class

# The Impact of Gender

## Experiment1: Conclusion

**Accuracy**

**Time**

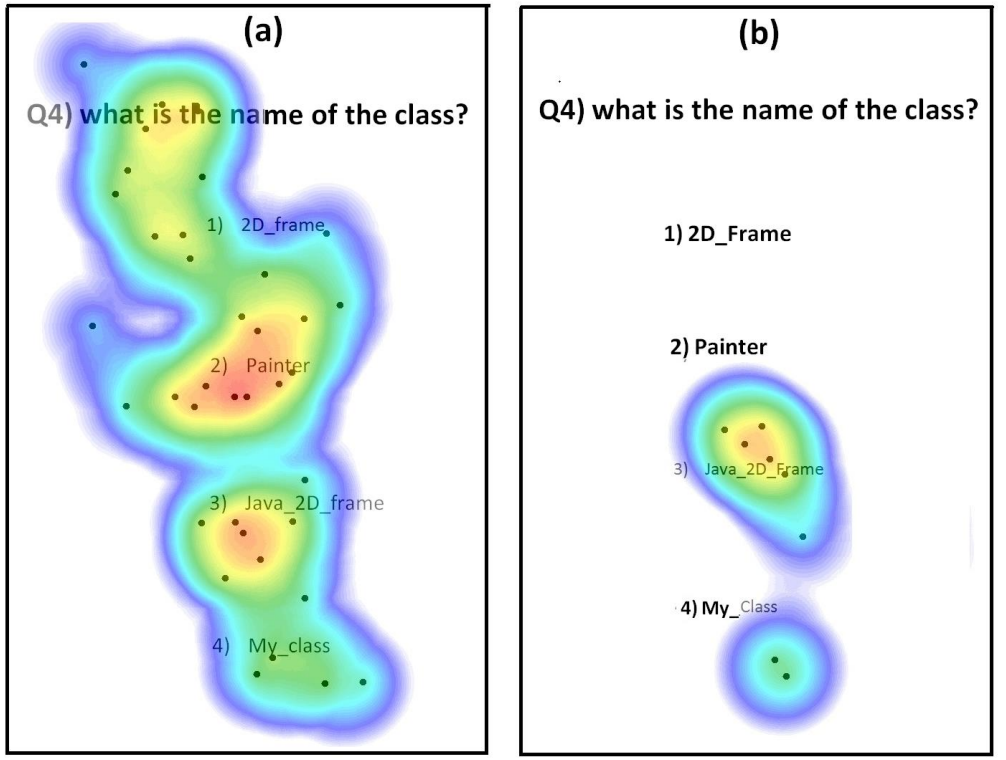


**Effort**

Men and women use different strategies to select the correct answer

# The Impact of Gender

## Lesson Learned


Different Attention Distribution

# The Impact of Gender

## Experiment 2: the Impact of SCEs

- Previous studies suggest that developers:
  - Avoid understanding the entire system
  - Focus on some parts

- Source code entities

Class name | Method name | Comments | Variables

# The Impact of Gender
## Experiment 2: the impact of SCEs

RQ3: What are the important source code entities (SCEs)

RQ4: Impact of gender on developers' viewing strategies

EMSE, 2014 & ICSM, 2012

# The Impact of Gender
# Experiment 2: Task

```
// This class is responsible to communicate with database
// with the help of JDBC driver                                    1

public class DataDB {                                              2
    Connection myc = null;
    Statement newS = null;
    ResultSet newSet = null;                                       3

    String myAdrs = "com.microsoft.jdbc.sqlserver.SQLServerDriver";
    String myPath = "jdbc:microsoft:sqlserver://localhost:1433;DatabaseName-MyDatabase";

    public void getMyTableData() {                                 4

        String newWord = "";

        myC = DriverManager.getConnection(dBUrl, "userid", "password");
        newS = dBConnection.createStatement();
        dBQuery = "select * from myTable";
        newSet = newS.executeQuery(newWord);

        while (newSet.next()) {
            System.out.print(newSet.getString(1) + ","
                    + dBResultSet.getString(2));

            System.out.print("," + newSet.getString(3) + "\n");

        }
    }

}
```

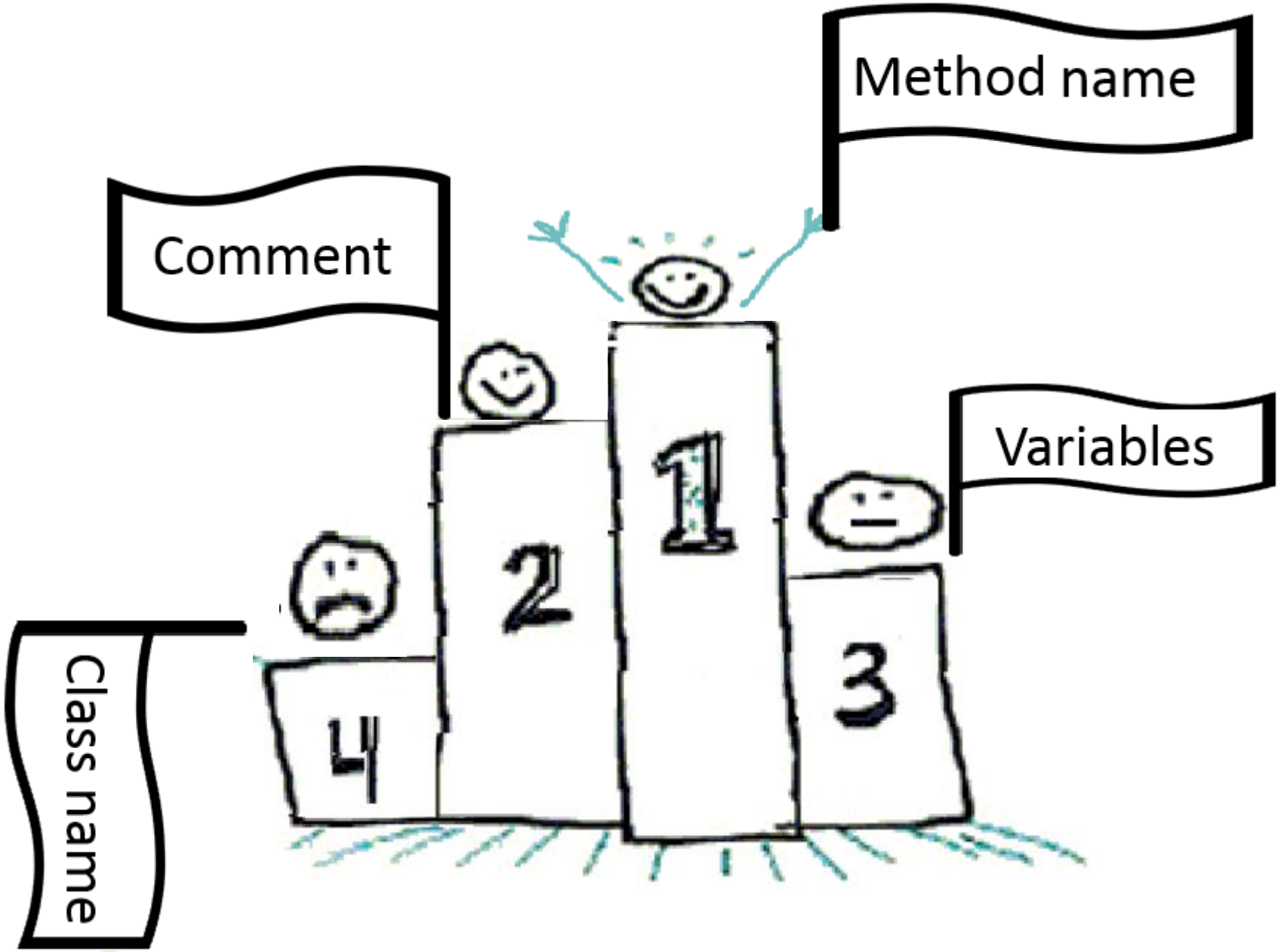
This class uses JDBC driver to connect to a database and get data from myTable

# The Impact of Representation Type

## Participants

Total: 24

| Academic Background | | | Gender | |
|---|---|---|---|---|
| **PhD** | **Masters** | **BSc** | **Male** | **Female** |
| 20 | 4 | 0 | 17 | 7 |


Method name
Comment
Variables
Class name
1
2
3
4

# The Impact of Gender

## Experiment 2: Viewing Strategies

```
// This class is responsible to communicate with database
// with the help of JDBC driver                                    1

public class DataDB {                                              2
    Connection myc = null;
    Statement newS = null;
    ResultSet newSet = null;                                       3

    String myAdrs = "com.microsoft.jdbc.sqlserver.SQLServerDriver";
    String myPath = "jdbc:microsoft:sqlserver://localhost:1433;DatabaseName-MyDatabase";

    public void getMyTableData() {                                 4

        String newWord = "";

        myC = DriverManager.getConnection(dBUrl, "userid", "password");
        newS = dBConnection.createStatement();
        dBQuery = "select * from myTable";
        newSet = newS.executeQuery(newWord);

        while (newSet.next()) {
            System.out.print(newSet.getString(1) + ","
                    + dBResultSet.getString(2));

            System.out.print("," + newSet.getString(3) + "\n");

        }
    }

}
```

This class uses JDBC driver to connect to a database and get data from myTable

# The Impact of Gender

## Experiment 2: Viewing Strategies

Path: 1,2,1,2,3,4

```
// This class is responsible to comm
// with the help of JDBC driver                    1

publi  clas  DataDB {                              2
    Connection myc = null;
    Statement newS = null;
    ResultSet newSet = null;                       3

    String myAdrs = "com.microsoft.j
    String myPath = "jdbc:microsoft:

    public void getMyTableData() {                 4
```

Total time: 1780

Path: 1,2,3,2,3,1,2,4,3,1,3,4,3

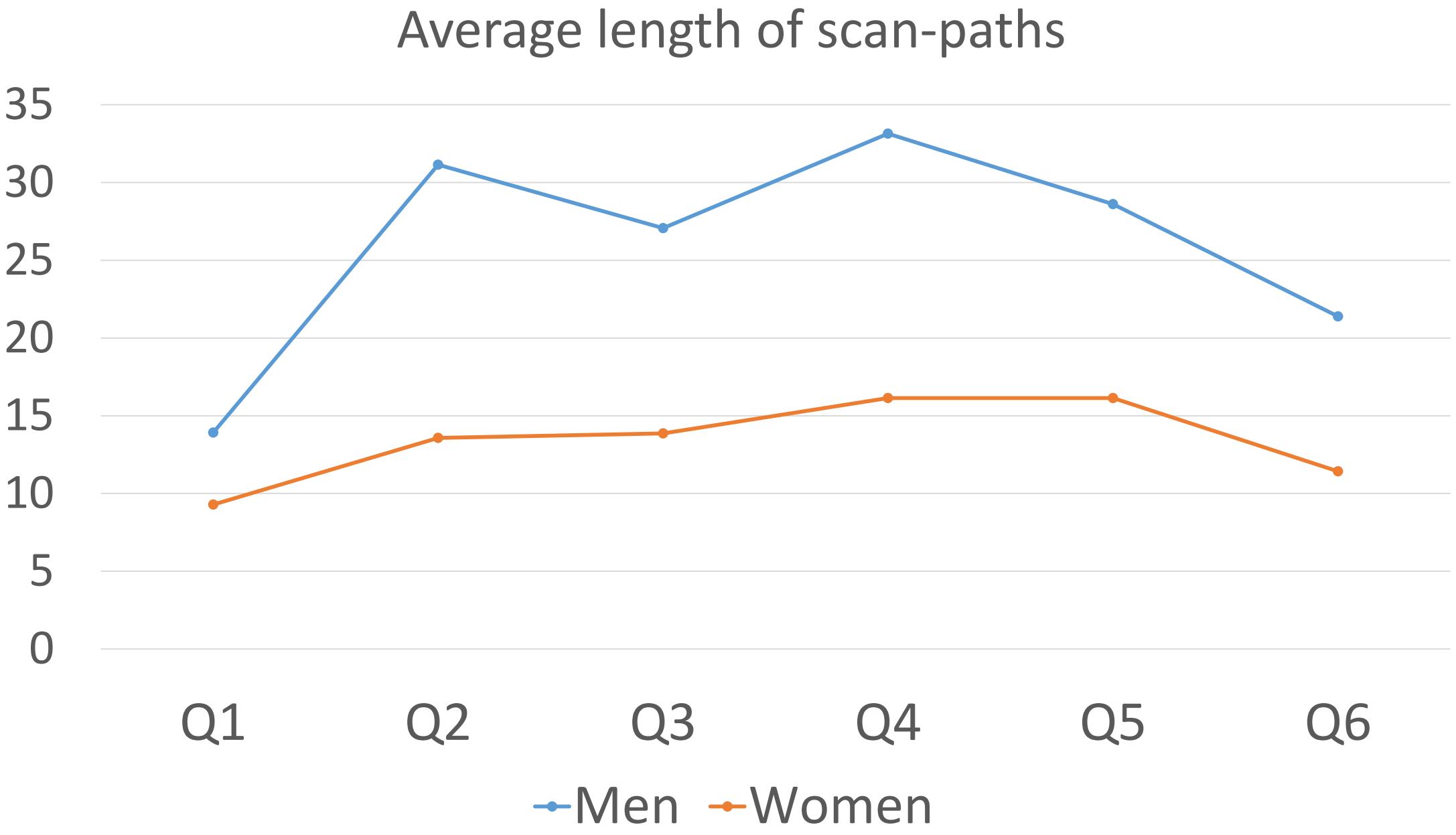
```
// This class is responsible to commu
// with the help of JDBC driver                    1

public class DataDB {                              2
    Connection myc = null;
    Statement newS = null;
    ResultSet newSet = null;                       3

    String myAdrs = "com.microsoft.jd
    String myPath = "jdbc:microsoft:s

    public void getMyTableData() {                 4
```

Total time: 1315


Average length of scan-paths
35
30
25
20
15
10
5
0
Q1
Q2
Q3
Q4
Q5
Q6
Men
Women

# The Impact of Gender
## Further Analysis

Method name is the most preferred type of SCEs.

```
// This class accesses all files of a directory and
// finds files with specific type.

public class FileUtil  {

    //create a FileFilter and override its accept-method
    FileFilter filefilter = new FileFilter() {

        public boolean accept(File file) {
            if (file.getName().endsWith(".csv")) {
                return true;
            }
            return false;
        }
    };

    public void listFilesMethod(String dir) {

        File directory = new File(dir);
```

```
// This class accesses all files of a directory and
// finds files with specific type.

public class FileUtil        {

    //create a FileFilter and override its accept-method
    FileFilter filefilter = new FileFilter() {

        public boolean accept(File file) {
            if (file.getName().endsWith(".csv")) {
                return true;
            }
            return false;
        }
    };

    public void listFilesMethod(String dir) {

        File directory = new File(dir);
```

# The Impact of Gender
## Further Analysis

- No difference between MiB and CiB
- Participants start looking at the **Bigger** SCE
- Method name, method name, method name …

# The Impact of Gender

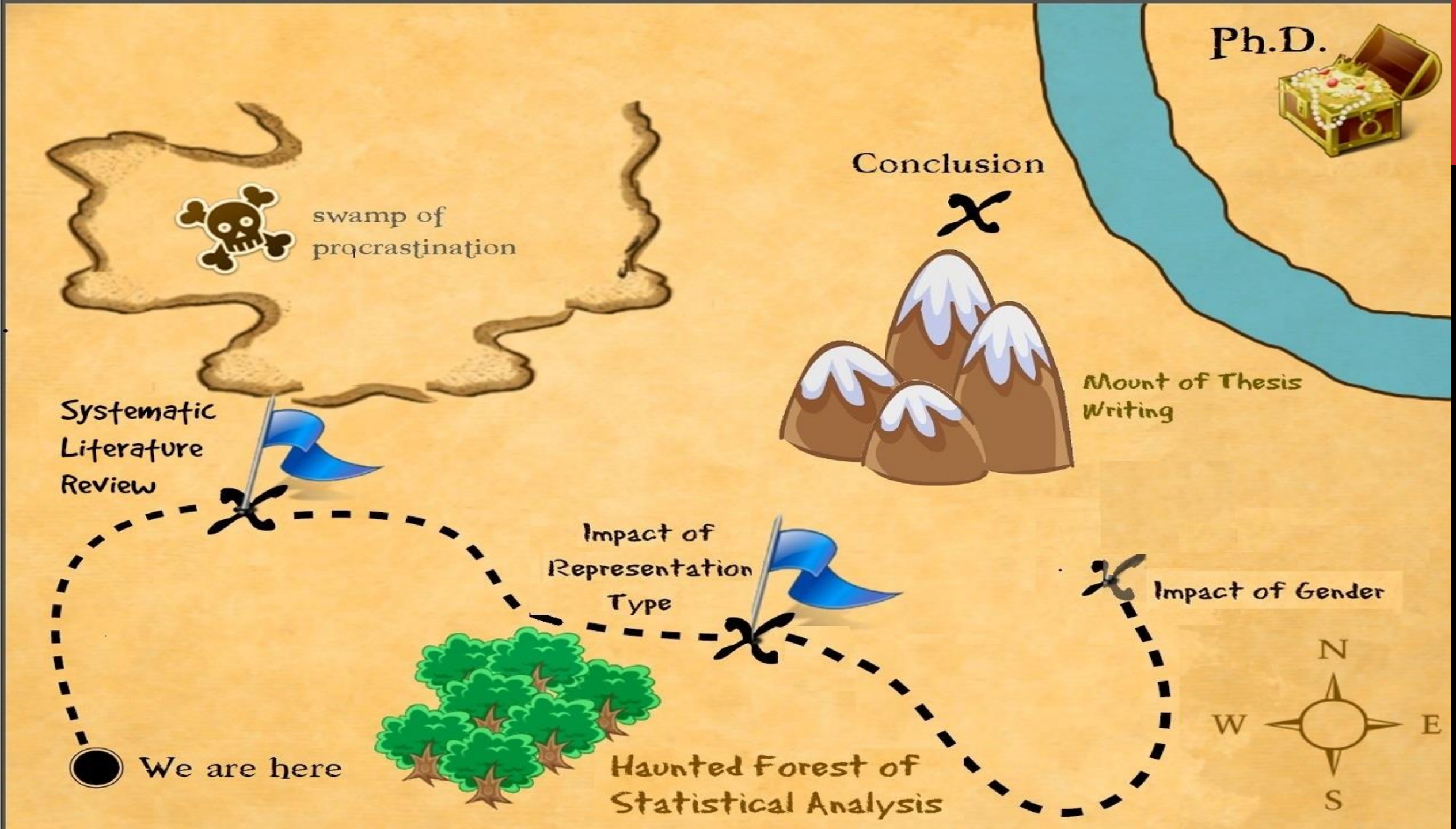
## Lesson Learned

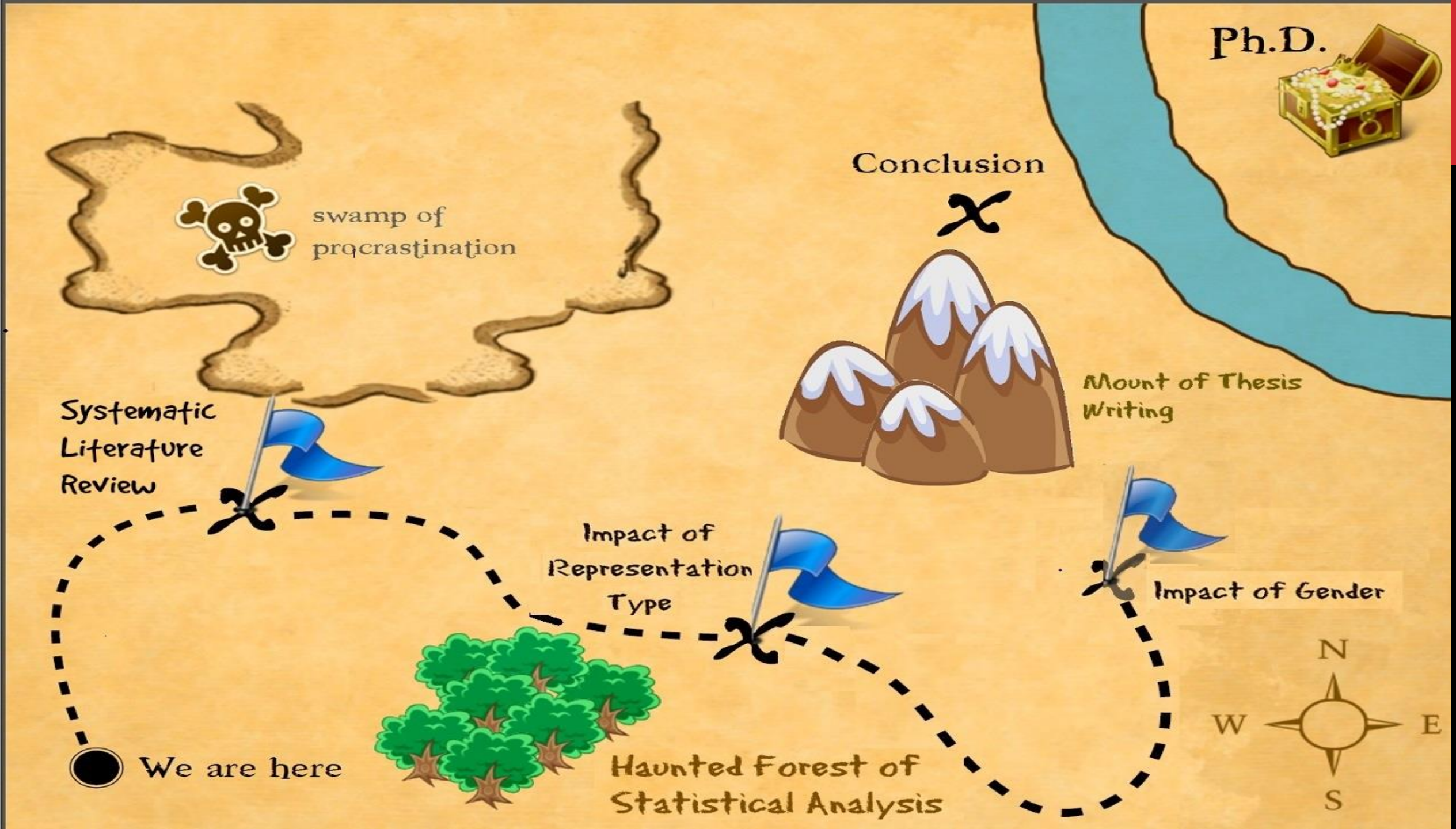
- Do not rush into adaptation

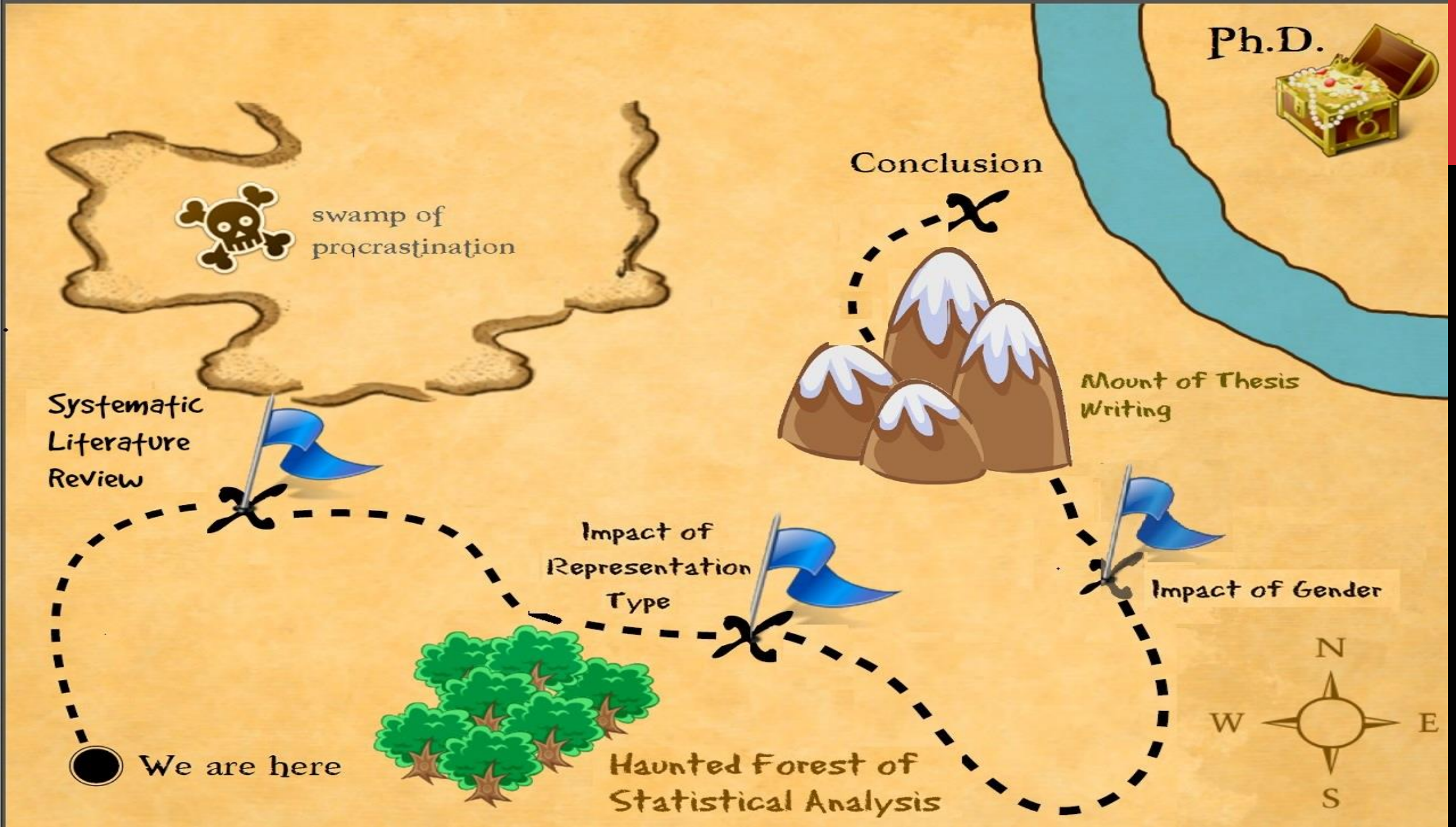
# The Impact of Gender
## Conclusion

Our findings raise two significant new open questions:

1. Are there differences between viewing strategies of male and female developers while performing program comprehension tasks and whether these strategies impact their efficiency and effectiveness or not?
2. To which extent do current programming environments support these strategies?


Ph.D.
Conclusion
swamp of
procrastination
Mount of Thesis
Writing
Systematic
Literature
Review
Impact of
Representation
Type
Impact of Gender
We are here
Haunted Forest of
Statistical Analysis
N
W
E
S


Ph.D.
Conclusion
swamp of
procrastination
Mount of Thesis
Writing
Systematic
Literature
Review
Impact of
Representation
Type
Impact of Gender
We are here
Haunted Forest of
Statistical Analysis
N
W
E
S


Ph.D.
Conclusion
swamp of
procrastination
Mount of Thesis
Writing
Systematic
Literature
Review
Impact of
Representation
Type
Impact of Gender
We are here
Haunted Forest of
Statistical Analysis
N
W
E
S

Vs.

Apple

---

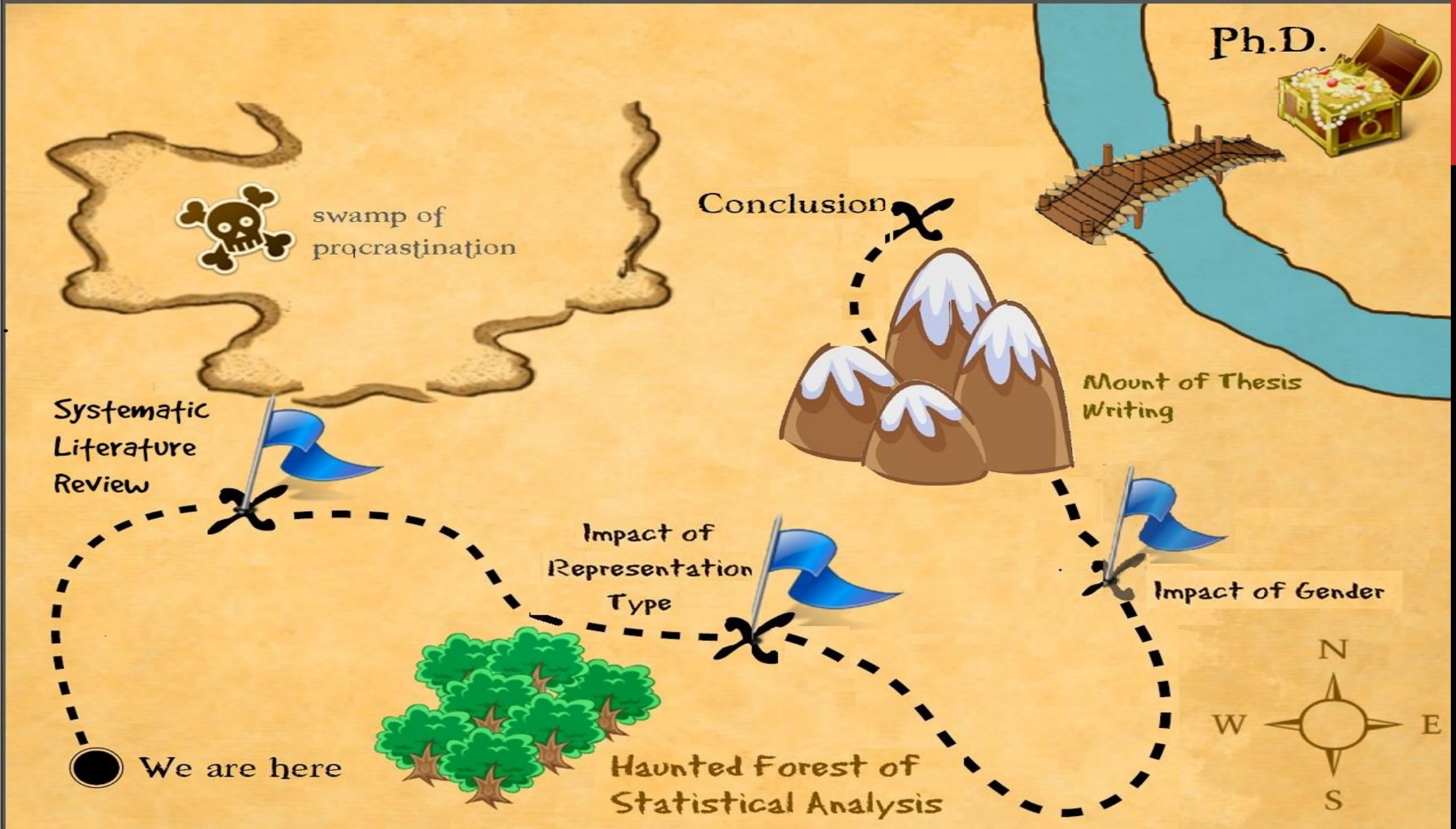
- Round
- Red or Yellow
- Edible fruit

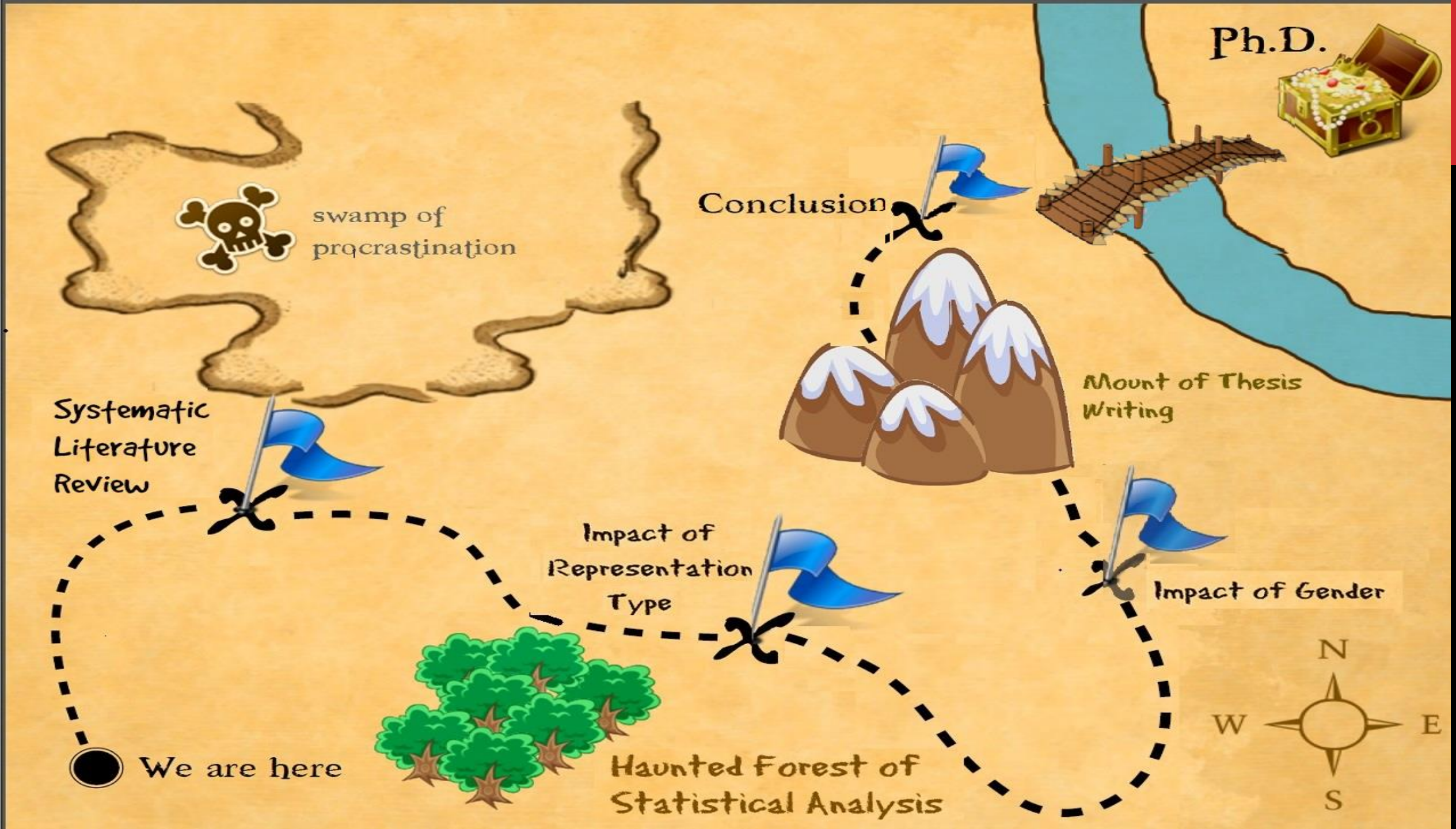

Short term

- Analyze the impact of layout on developers' performance
- Scan-path and AOI analysis
- Image processing approaches to analyze heatmaps

Long term

- Study developers' viewing strategies and reading behaviour
- Analyze the applicability of our gender differences findings


Ph.D.
swamp of procrastination
Conclusion
Mount of Thesis Writing
Systematic Literature Review
Impact of Representation Type
Impact of Gender
We are here
Haunted Forest of Statistical Analysis
N
W
E
S


Ph.D.
swamp of procrastination
Conclusion
Mount of Thesis Writing
Systematic Literature Review
Impact of Representation Type
Impact of Gender
We are here
Haunted Forest of Statistical Analysis
N
W
E
S

## Awards

1. Le Fonds de recherche du Québec, Nature et technologies (FRQNT)

## Journal Papers

1. Taupe: Visualizing and Analysing Eye-tracking Data**. SCP, 2011**
2. An empirical study on the importance of source code entities for requirements traceability. **EMSE, 2014**
3. Systematic Literature Review on the Use of Eye-tracking Technique in Software Engineering. Under revision, **IST, 2014**

## Conference Papers

1. Women & Men: Different but Equal: A Study on the Impact of Identifiers on Source Code Understanding. **ICPC, 2012**
2. Professional Status or Expertise for UML Class Diagram Comprehension: An Empirical Study Systematic. **ICPC 2012**
3. An Empirical Study on Requirements Traceability Using Eye-Tracking. **ICSM, 2012**
4. An Empirical Study on the Efficiency of Graphical vs. Textual in Requirements Comprehension. **ICPC, 2013**

Vs.

Apple

---

- Round
- Red or Yellow
- Edible fruit