BIODIV Advanced Courses

Introduction to Programming with R

Pedro Tarroso (ptarroso@cibio.up.pt)

Antigoni Kaliontzopoulou
Jesús Muñoz Pajares
Urtzi Enriquez-Urzelai

4-8 November, 2019

InBIO / CIBIO

Overview of the course

- Day 1 - Monday
  1. What is R? Installation (R and R-studio), main structure, seeking help
  2. Assignment, variable modes and operators
  3. Functions and packages (and seeking more help)
  4. Code style guidelines
- Day 2 - Tuesday
  1. Data types I: vectors, matrices and arrays
  2. Data type II: data frames, lists and other data types
  3. Data import-export, workspaces and directories
  4. Practice session: bring and load your own data
- Day 3 - Wednesday
  1. Data indexing
  2. Reviewing and tabulating your data
  3. String Manipulation
  4. Plotting
  5. Practice session: plots, indexing and data review
- Day 4 - Thursday
  1. Control functions I: conditionals
  2. Control functions II: loops
  3. Practice session: control functions
- Day 5 - Friday
  1. Some advanced features: build your functions
  2. Practice session: function writing
  3. Showcase: statistical analysis of biological data

Overview of the course

At the end of this course, you will be able to:

- Identify different components of code (variables, operators, functions, etc)
- Read and interpret R scripts written by others
- Write code in R programming language to perform multiple tasks
- Understand why you should use R to do your analyses
- Know basic functions in R to perform simple tasks
- Identify tasks where automation is useful
- Build and control the asthetics of plots
- Learn to solve a problem with programming strategy in mind

What is programming?

What is programming?

Programming is to create a list of instructions defining a set of operations to be executed by a computer to achieve a result

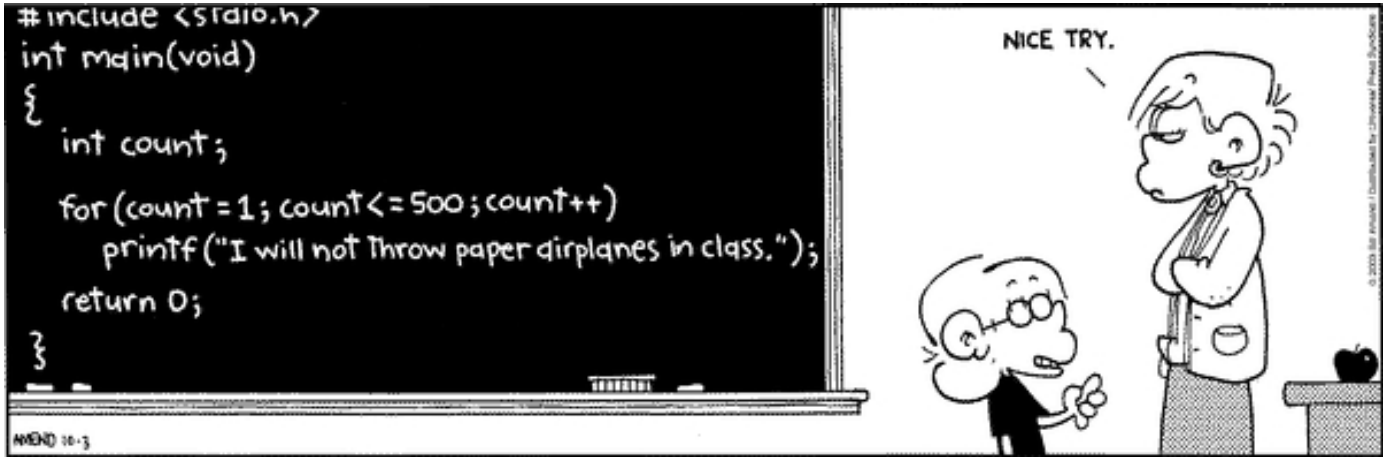
What is programming?

Programming is to create a list of instructions defining a set of operations to be executed by a computer to achieve a result

Programming is a creative process to instruct a machine on how to do a task

Why programming?

- Save time by automating repetitive tasks
- A single script can be applied to multiple datasets
- Maintain and improve the methods
- A permanent memory of the methods used
- Deal with big datasets
- Use methods from different areas
- Can be fun! (It is a creative process after all)



How to program?

How to program?

“

‘Where shall I begin, please your Majesty?’ he asked.

‘Begin at the beginning,’ the King said, gravely, ‘and go on till you come to the end: then stop.’

”

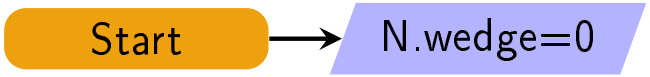
–Lewis Carrol - Alice’s Adventures in Wonderland

How to program?

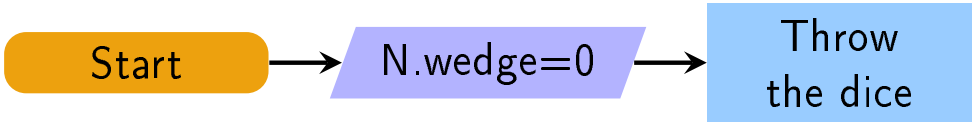
How to program?

Start

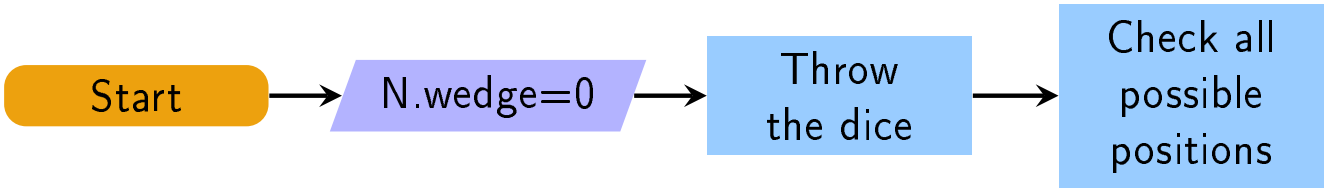
How to program?



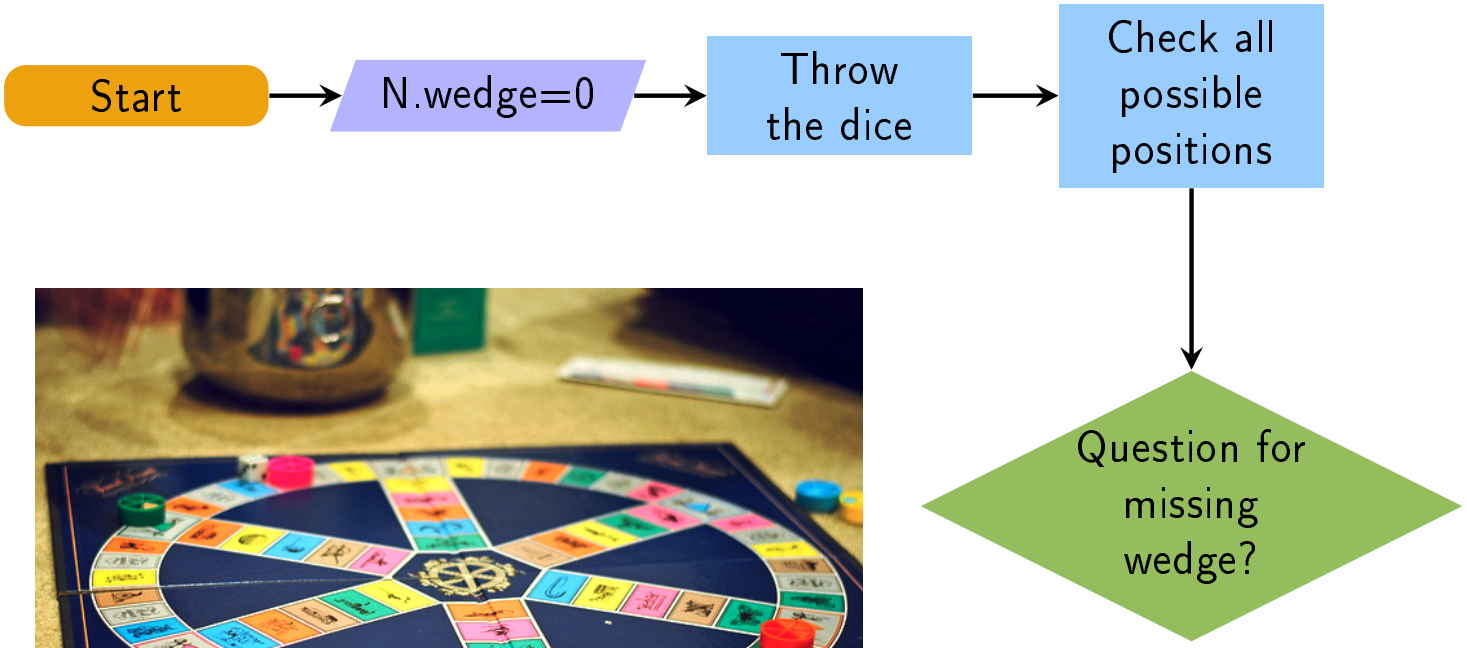
How to program?



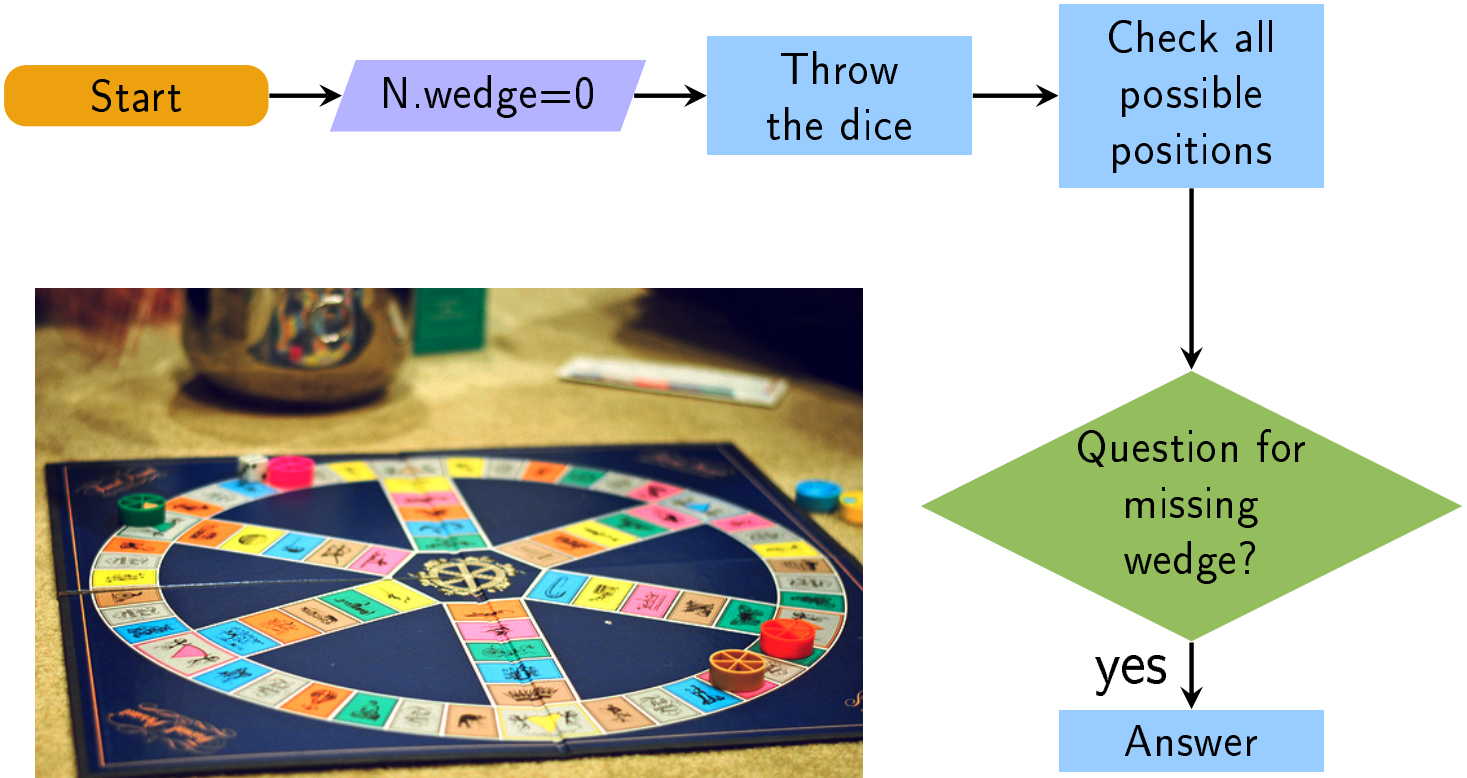
How to program?



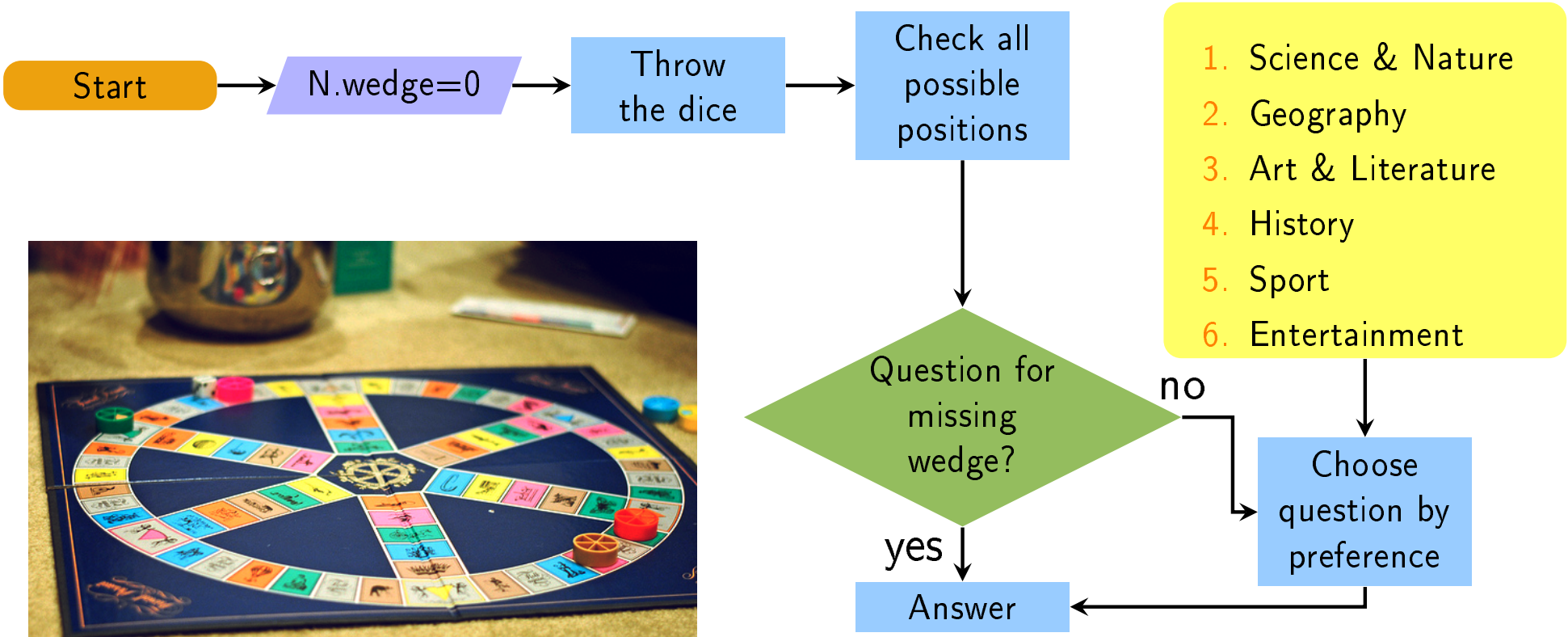
How to program?



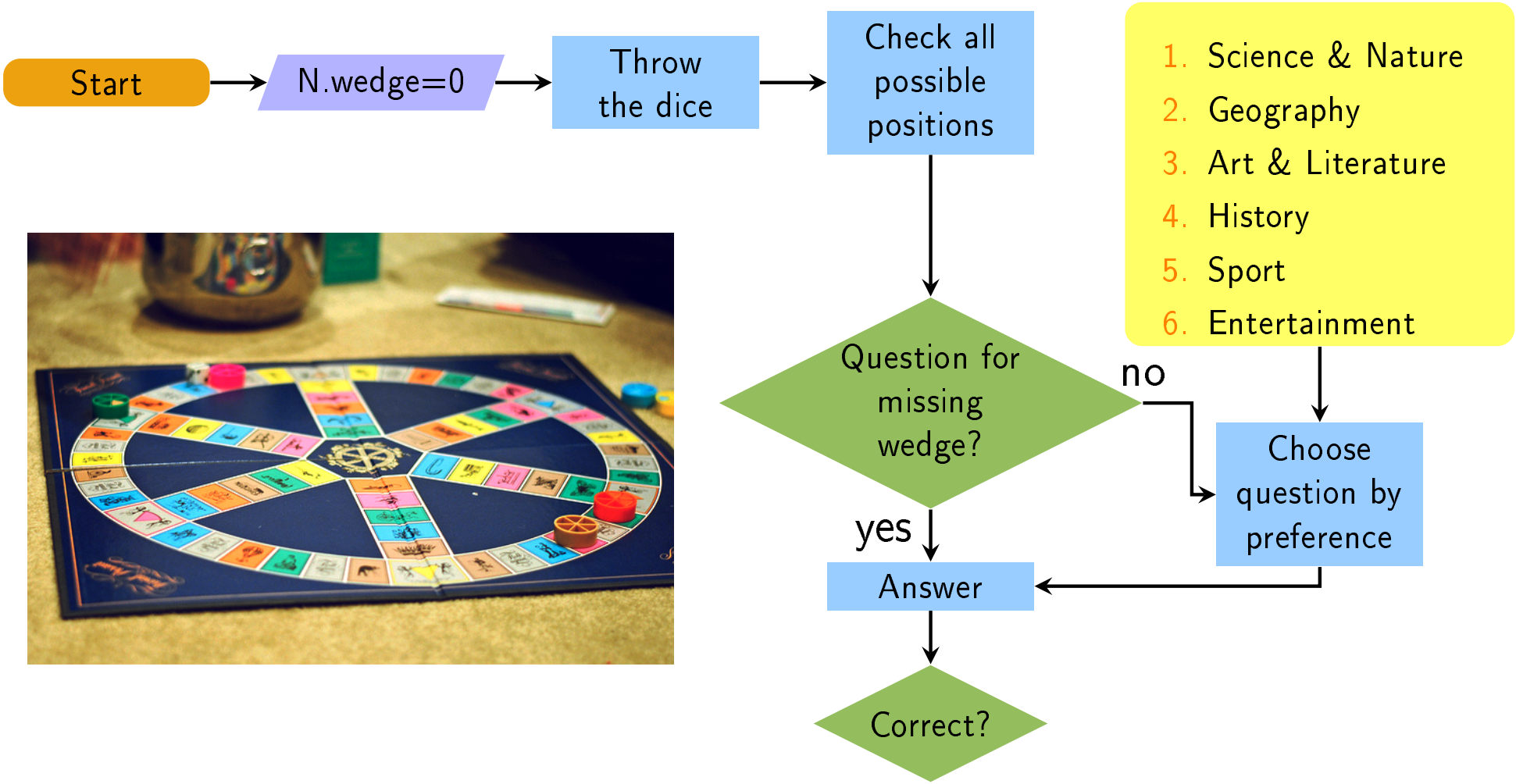
How to program?



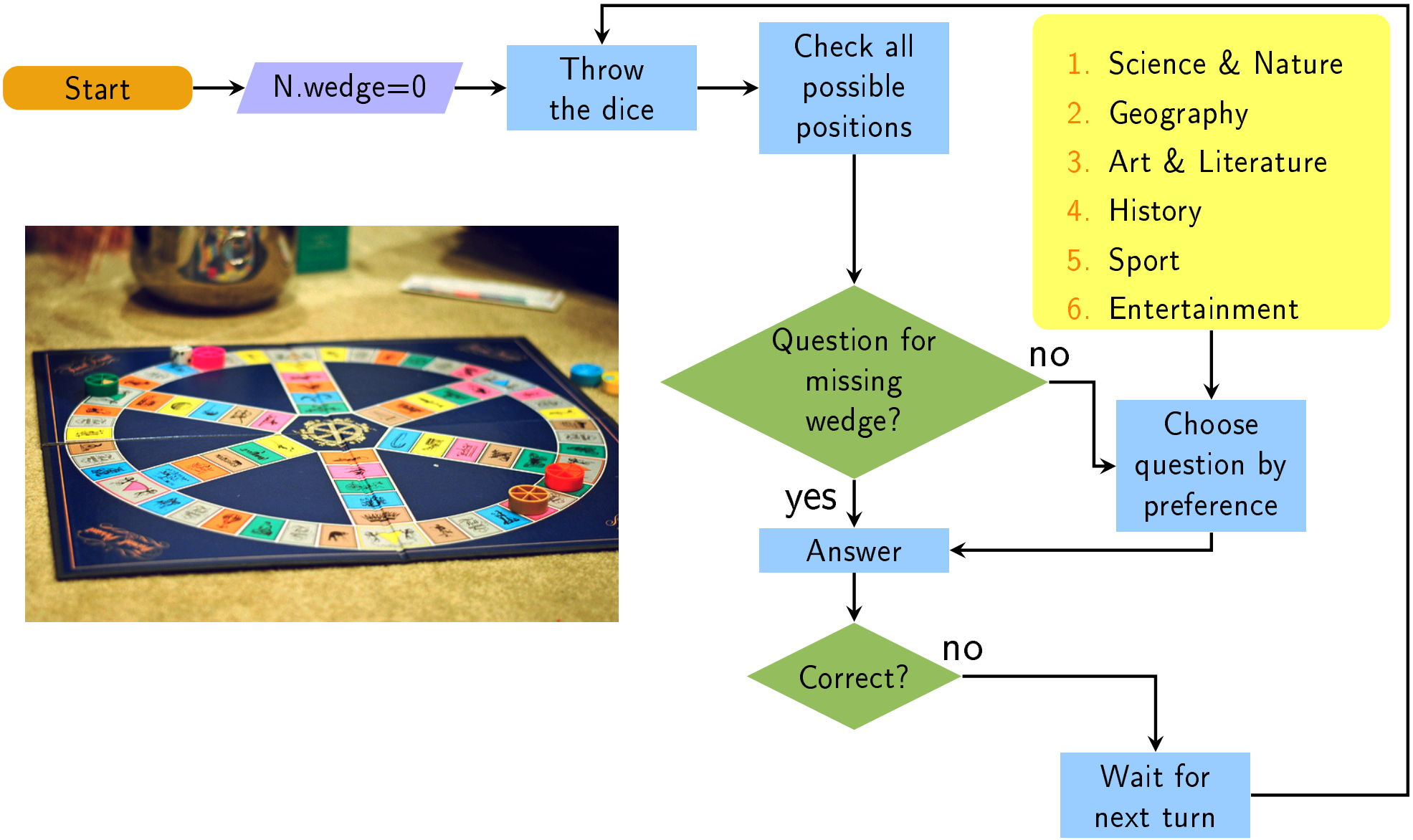
How to program?



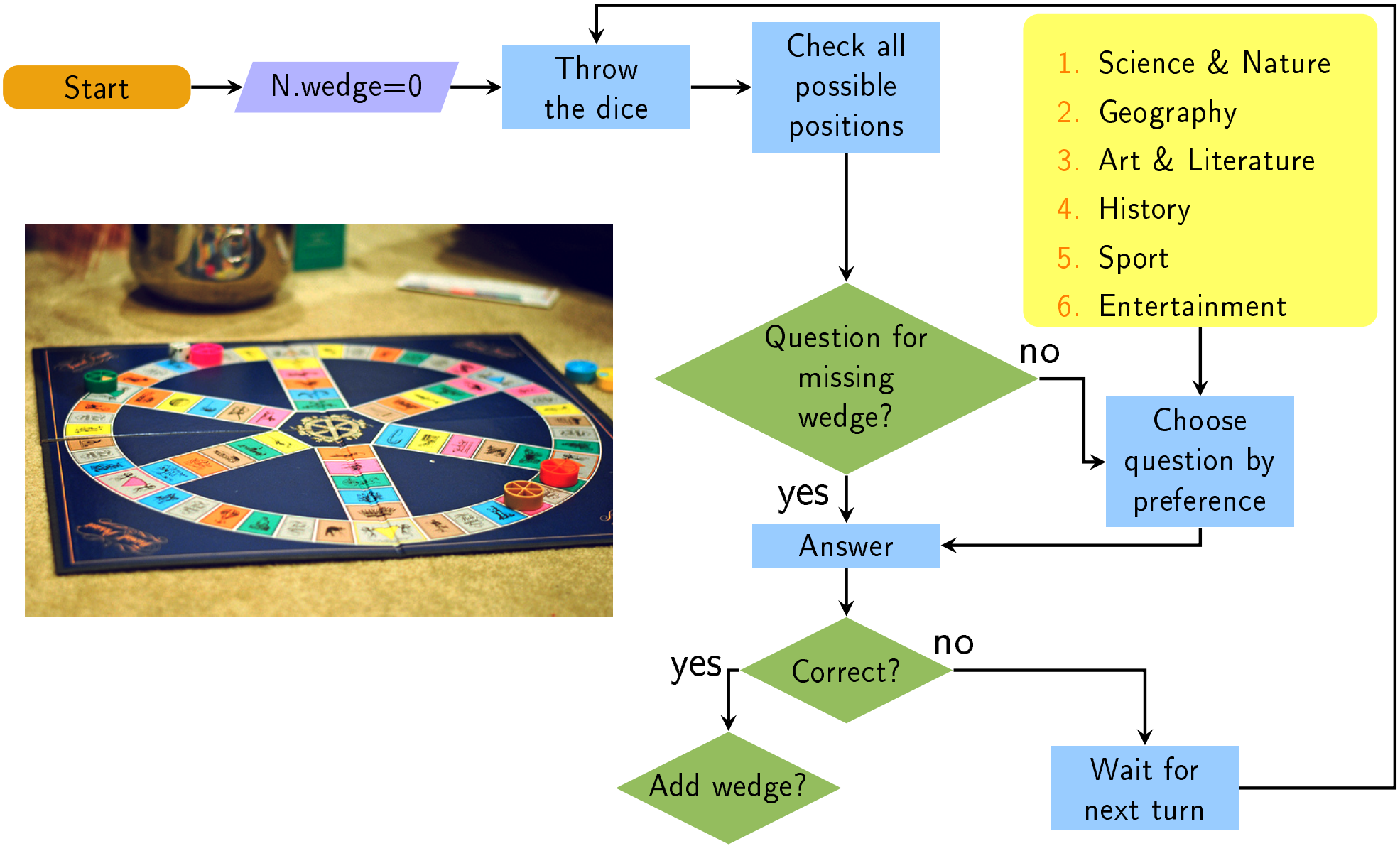
How to program?



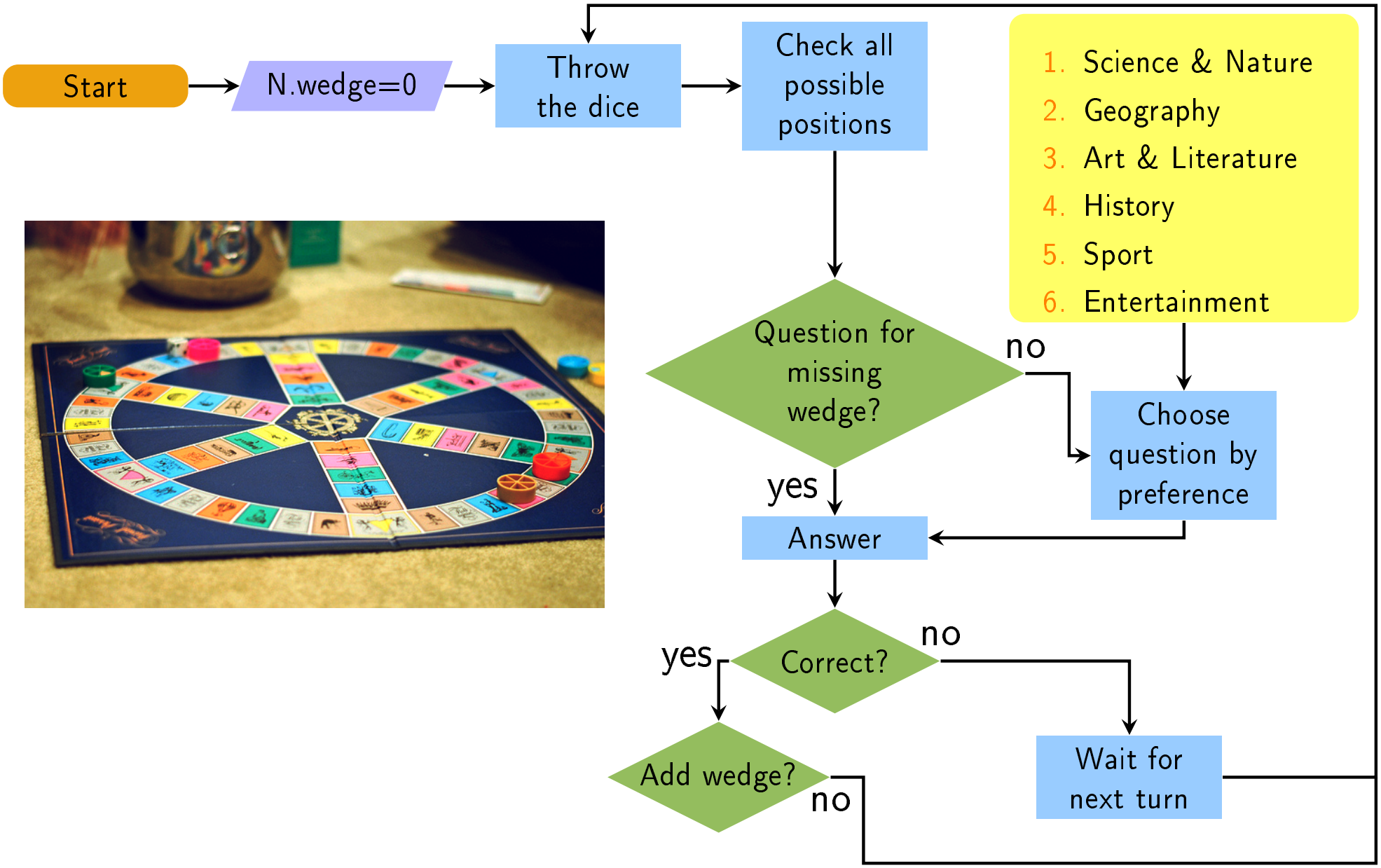
How to program?



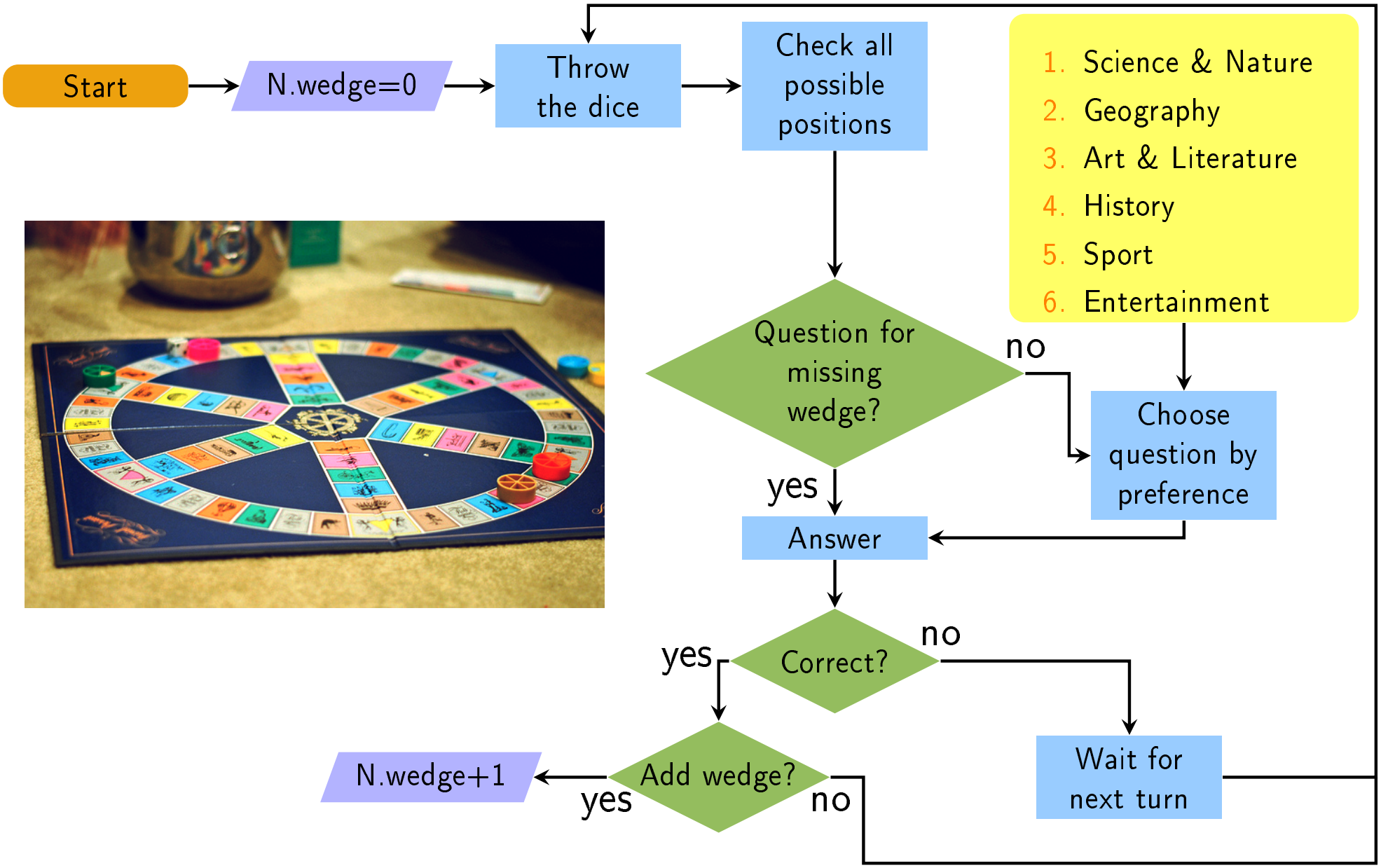
How to program?



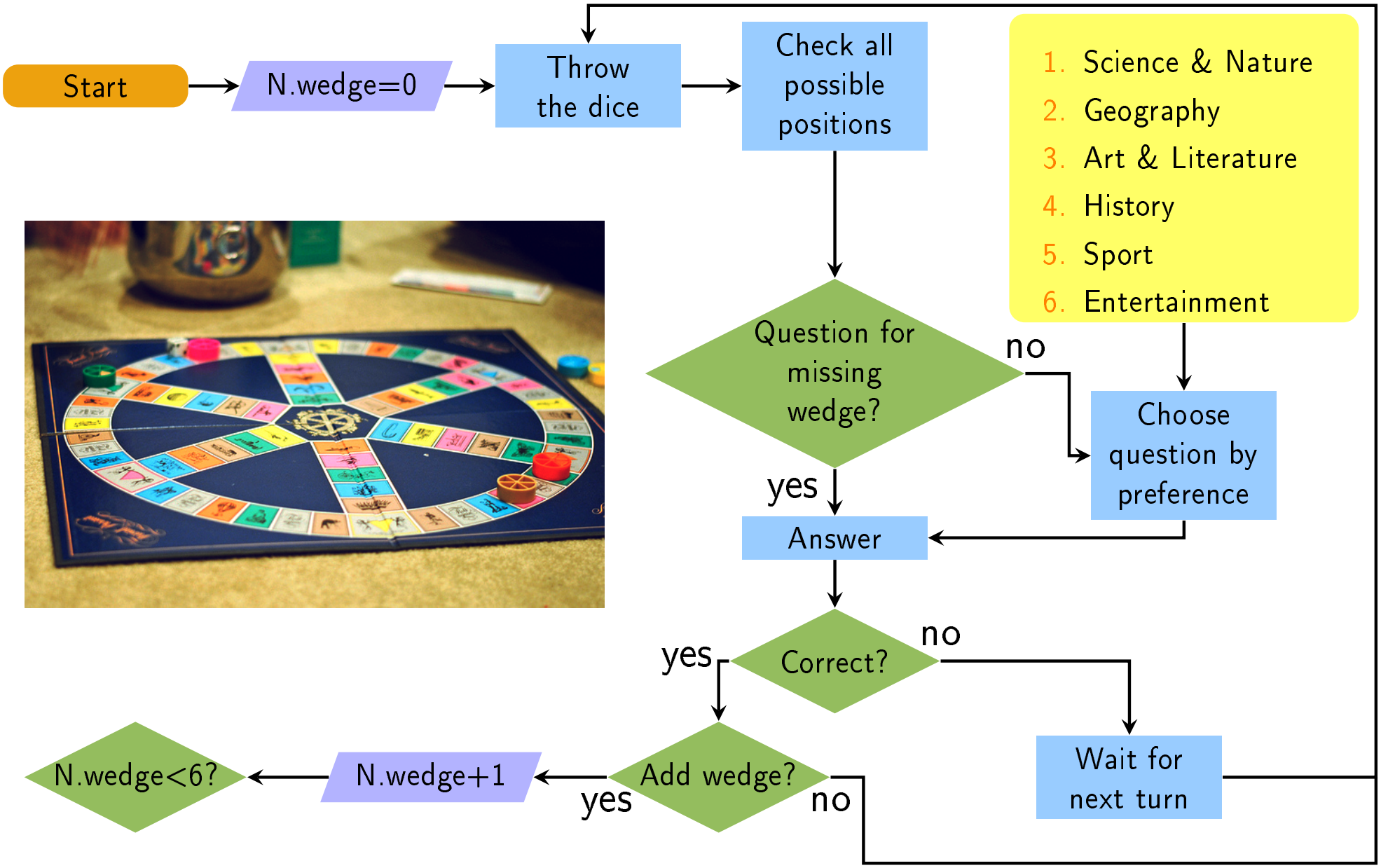
How to program?



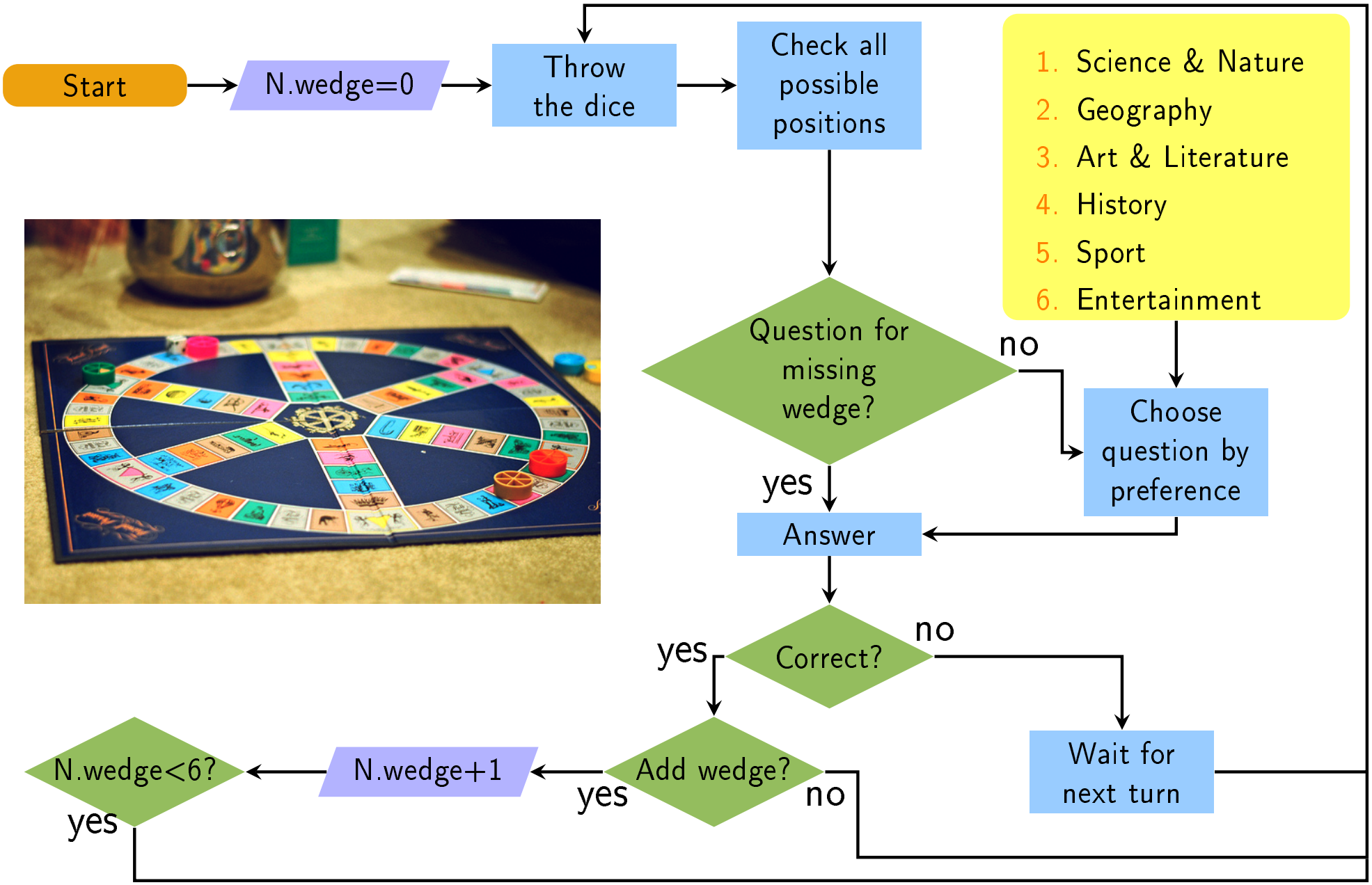
How to program?



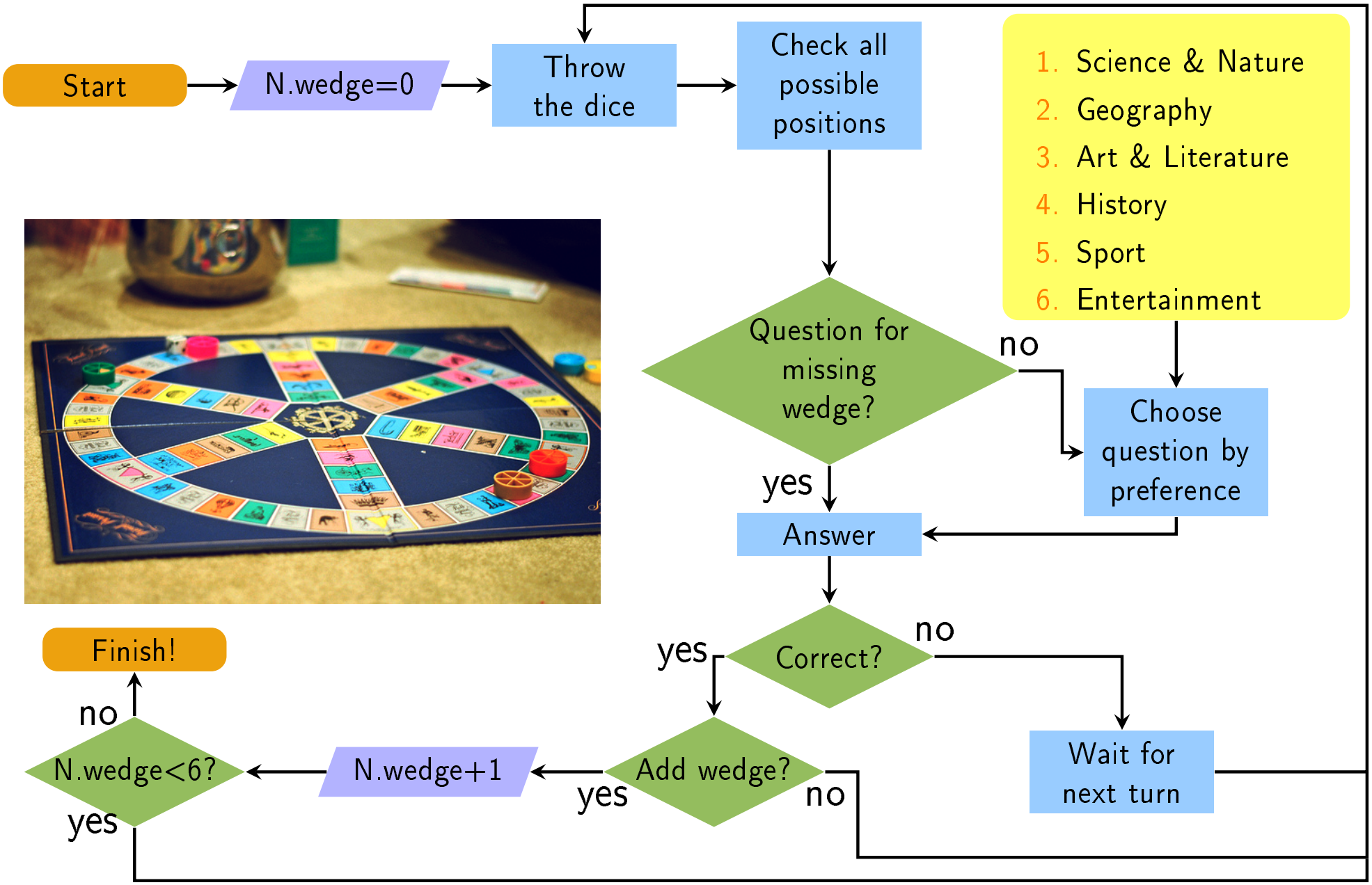
How to program?



How to program?



How to program?



How to program?

Sample	Length
1	3.4
2	2.7
3	3.2
4	3.7
5	2.9

What is the average length of the individuals?

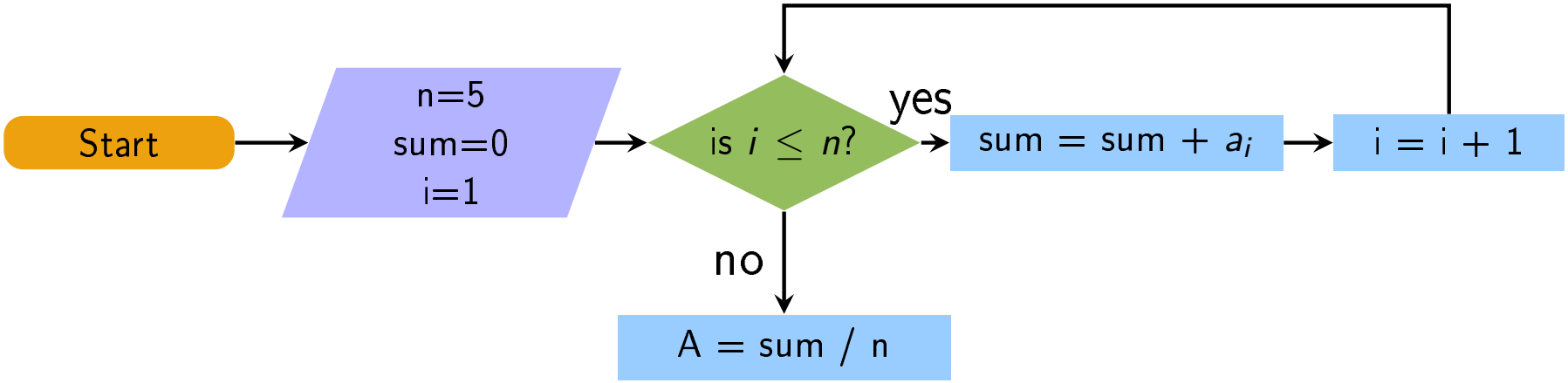
$$A = \frac{1}{n} \sum_{i=1}^{n} a_i$$

How to program?

Sample	Length
1	3.4
2	2.7
3	3.2
4	3.7
5	2.9

What is the average length of the individuals?

$$A = \frac{1}{n} \sum_{i=1}^{n} a_i$$

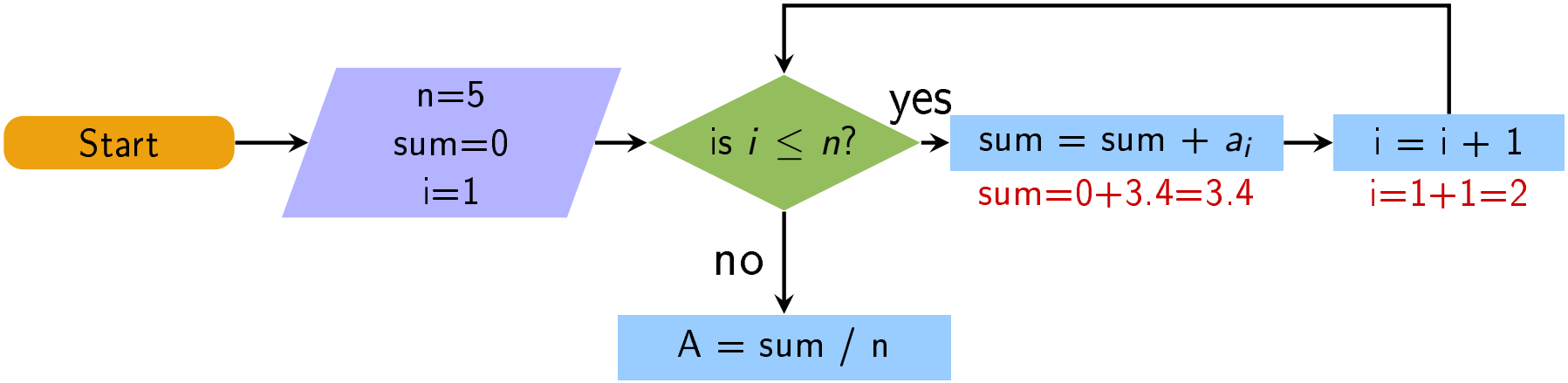


How to program?

Sample	Length
1	3.4
2	2.7
3	3.2
4	3.7
5	2.9

What is the average length of the individuals?

$$A = \frac{1}{n} \sum_{i=1}^{n} a_i$$

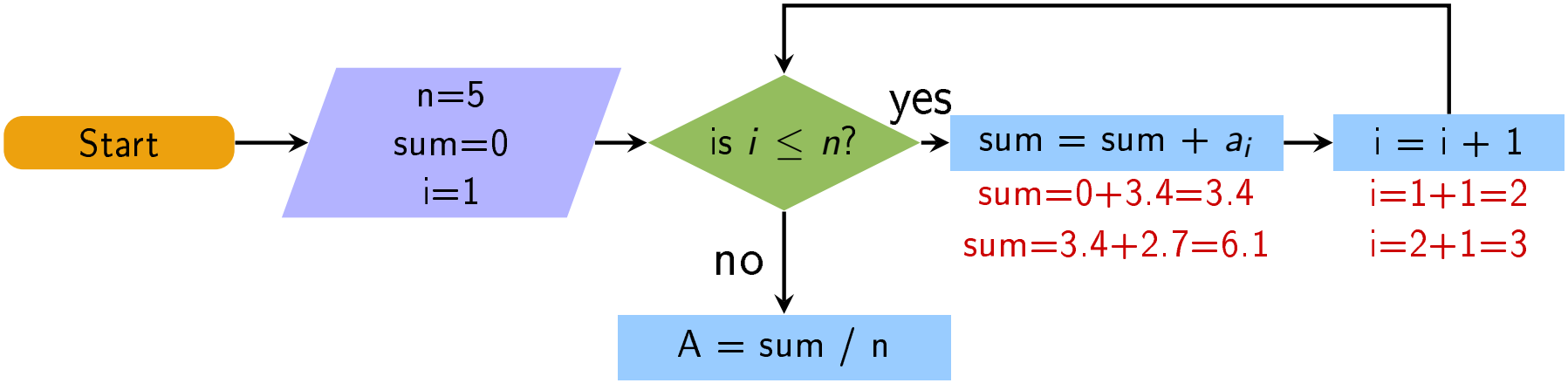


How to program?

Sample	Length
1	3.4
2	2.7
3	3.2
4	3.7
5	2.9

What is the average length of the individuals?

$$A = \frac{1}{n} \sum_{i=1}^{n} a_i$$

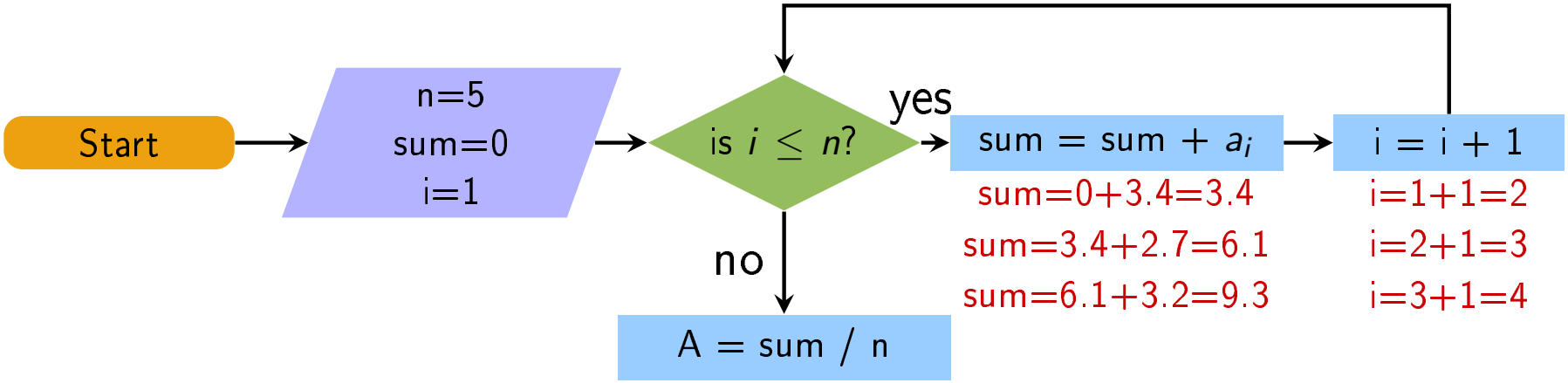


How to program?

Sample	Length
1	3.4
2	2.7
3	3.2
4	3.7
5	2.9

What is the average length of the individuals?

$$A = \frac{1}{n} \sum_{i=1}^{n} a_i$$

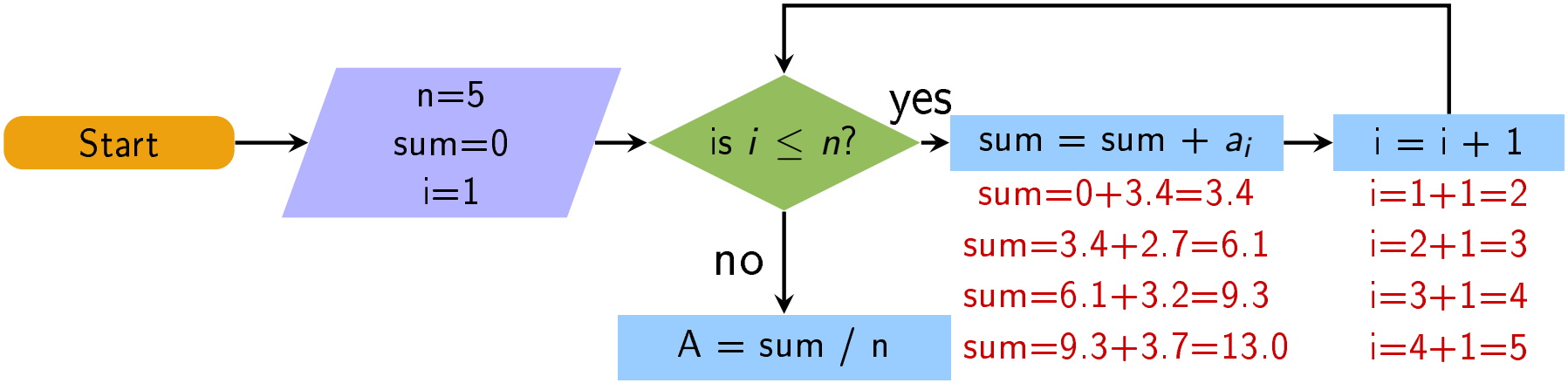


How to program?

Sample	Length
1	3.4
2	2.7
3	3.2
4	3.7
5	2.9

What is the average length of the individuals?

$$A = \frac{1}{n} \sum_{i=1}^{n} a_i$$

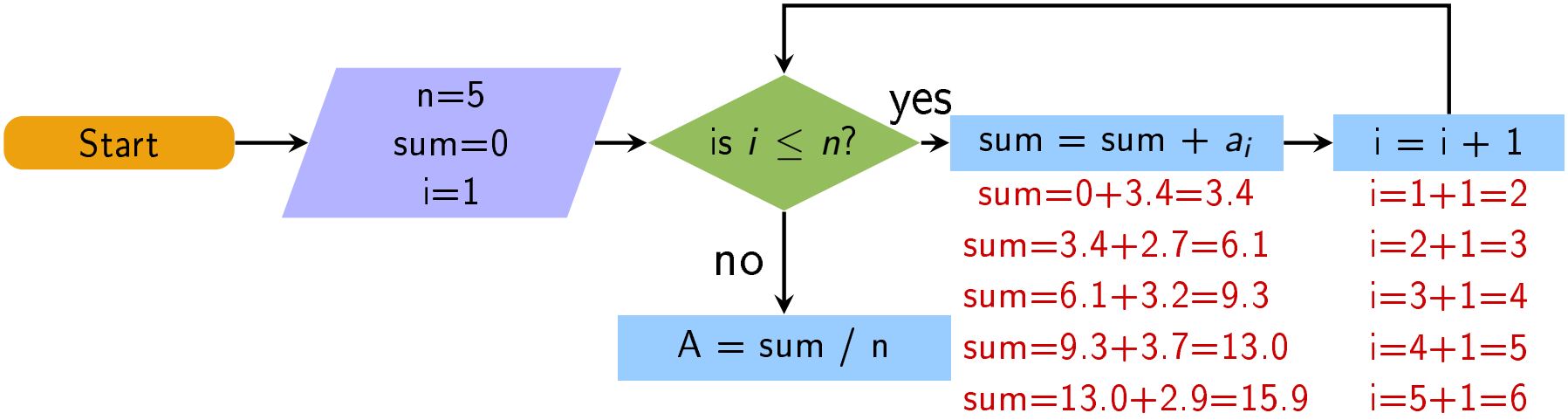


How to program?

Sample	Length
1	3.4
2	2.7
3	3.2
4	3.7
5	2.9

What is the average length of the individuals?

$$A = \frac{1}{n} \sum_{i=1}^{n} a_i$$

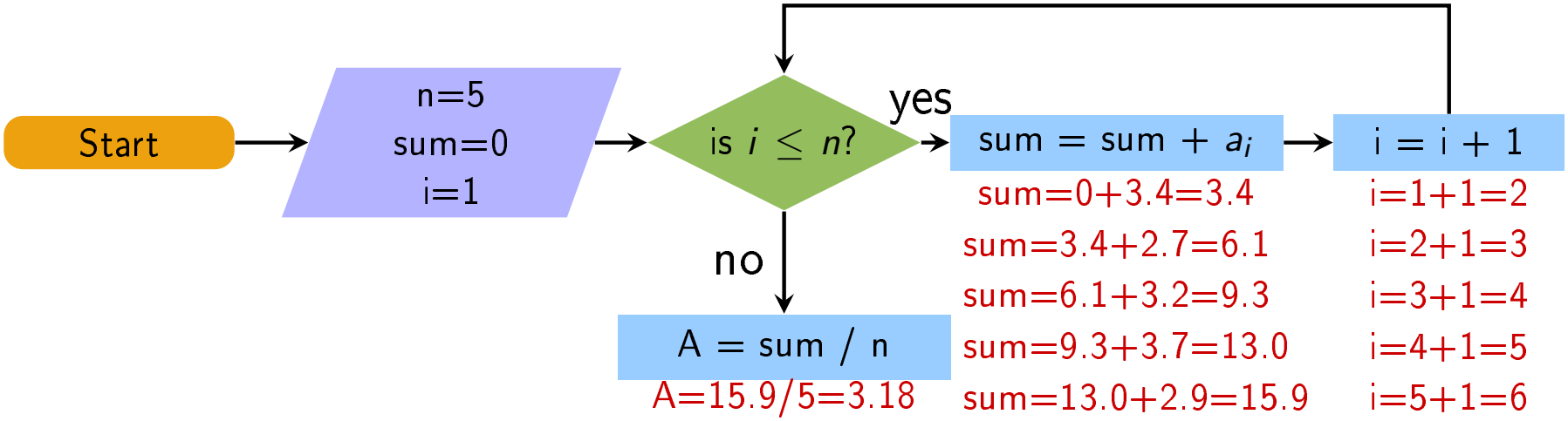


How to program?

Sample	Length
1	3.4
2	2.7
3	3.2
4	3.7
5	2.9

What is the average length of the individuals?

$$A = \frac{1}{n} \sum_{i=1}^{n} a_i$$

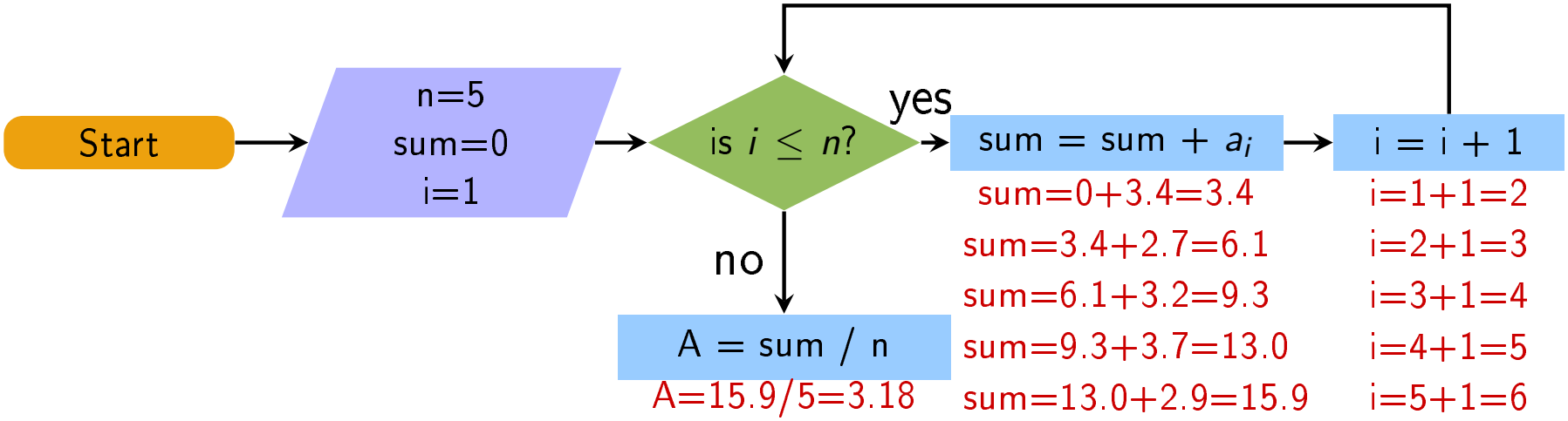


How to program?

Sample	Length
1	3.4
2	2.7
3	3.2
4	3.7
5	2.9

What is the average length of the individuals?

$$A = \frac{1}{n}\sum_{i=1}^{n} a_i$$



Why is R so easy to learn?

mean(Lengths)

How to program?

data.txt

```
12.1
12.5
11.2
9.4
15.8
2.5
6.8
14.2
13.6
11.4
12.6
11.8
11.1
```

How to program?

data.txt

12.1
12.5
11.2
9.4
15.8
2.5
6.8
14.2
13.6
11.4
12.6
11.8
11.1

C++

```cpp
#include <iostream>
#include <vector>
#include <fstream>
#include <string>
using namespace std;
int main()
{
    string txtfile;
    double dt;
    int n;
    vector<double> vec;
    txtfile = "data.txt";
    ifstream myfile(txtfile.c_str());
    if (myfile.is_open())
    {
        while (myfile >> dt)
         vec.push_back(dt);
    }
    myfile.close();
    n = vec.size();
    for (int i = 0; i < n; i++)
        cout << vec[i] << "␣";
        cout << "\n";
    }
```

Python

```python
myfile = 'data.txt'
mystream = open(myfile, 'r')
mydata = []
for line in mystream:
    mydata.append(float(line))

mystream.close()
print(mydata)
```

R

```r
myfile <- "data.txt"
mydata <- read.table(myfile)
print(mydata[,1])
```

Why is it so difficult to start?

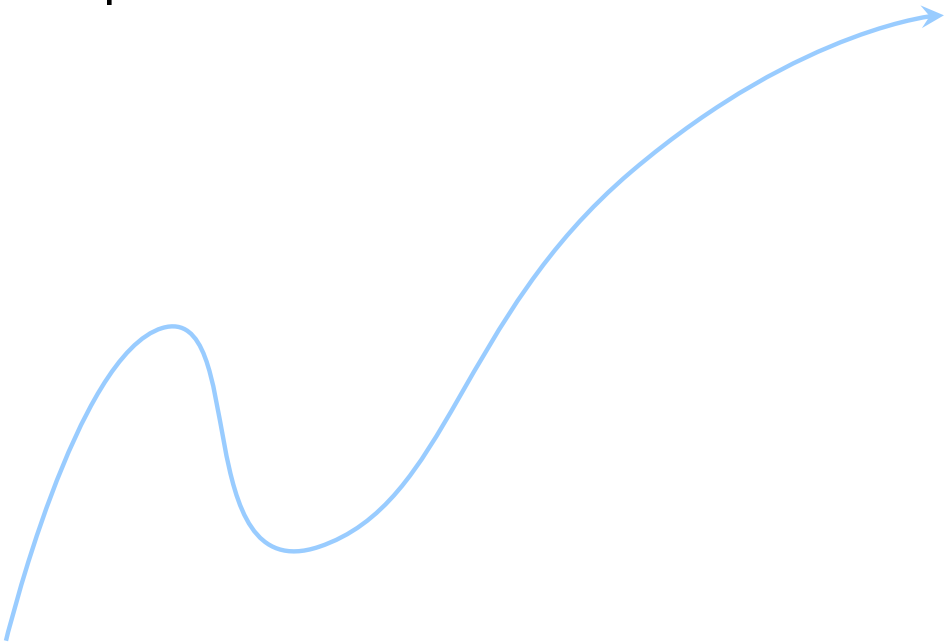
Learning a new language is never easy

- develop some programming intuition
- master some key concepts of the language to communicate with the computer
- programming logic often deviates from everyday common sense
- easy to give up when other more user-friendly tools are available
- difficult to express in code the idea of a solution to a problem

Why is it so difficult to start?

Learning a new language is never easy

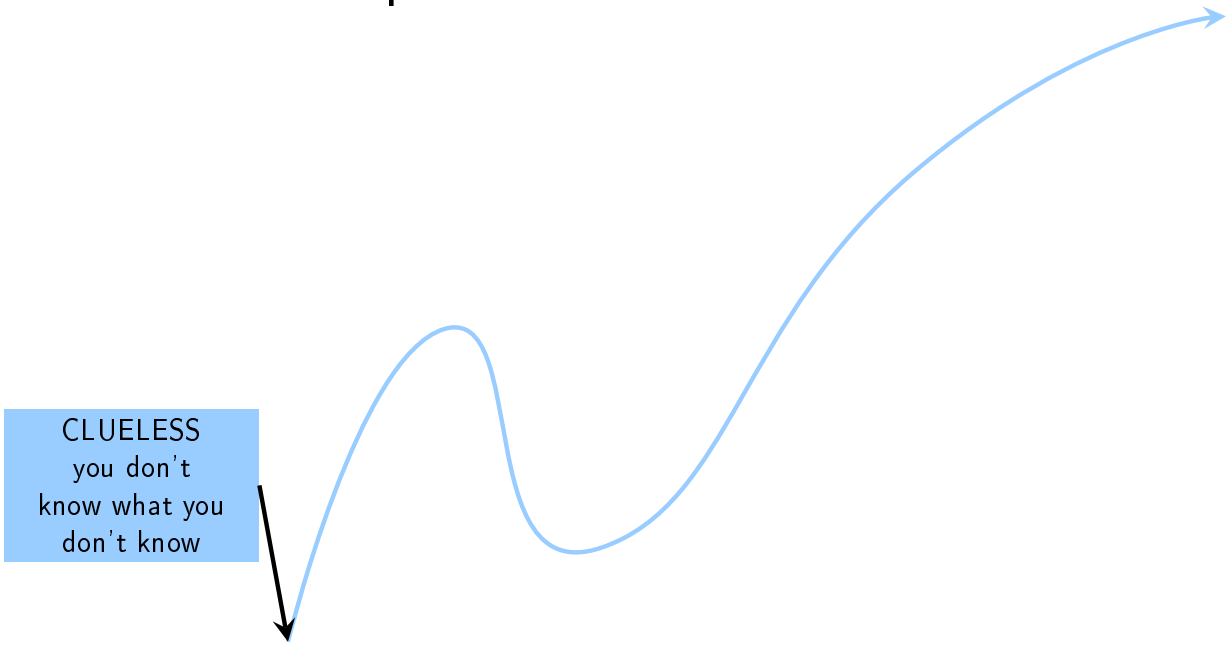
- develop some programming intuition
- master some key concepts of the language to communicate with the computer
- programming logic often deviates from everyday common sense
- easy to give up when other more user-friendly tools are available
- difficult to express in code the idea of a solution to a problem



Why is it so difficult to start?

Learning a new language is never easy

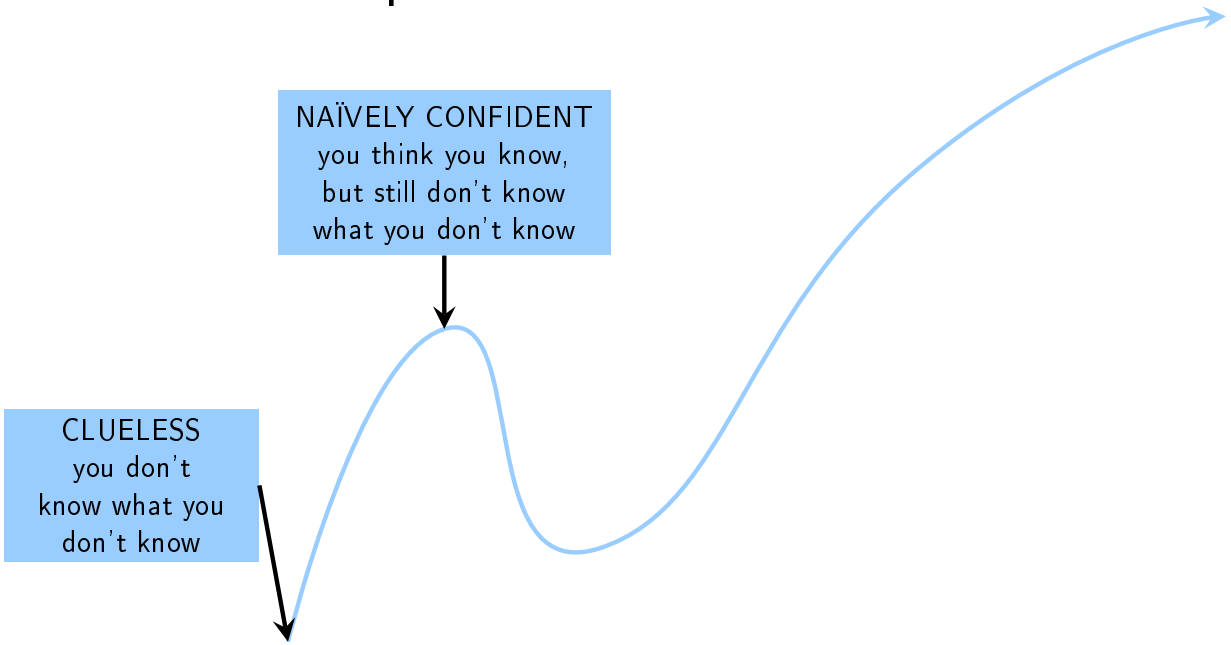
- develop some programming intuition
- master some key concepts of the language to communicate with the computer
- programming logic often deviates from everyday common sense
- easy to give up when other more user-friendly tools are available
- difficult to express in code the idea of a solution to a problem



Why is it so difficult to start?

Learning a new language is never easy

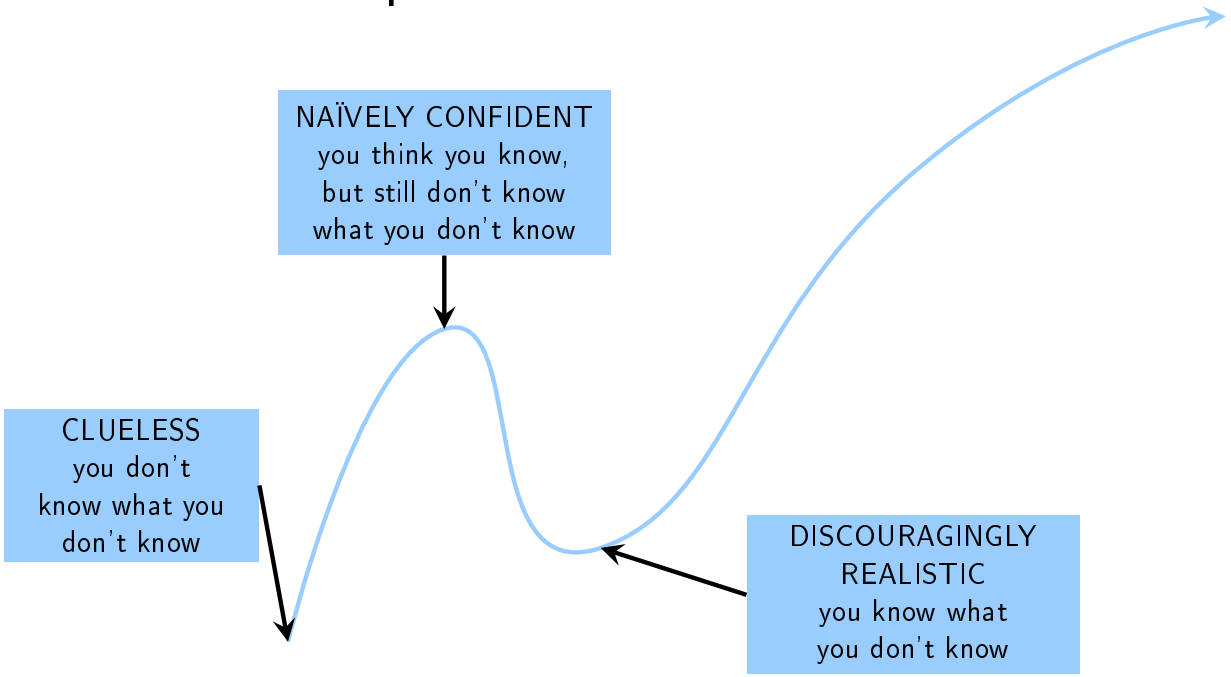
- develop some programming intuition
- master some key concepts of the language to communicate with the computer
- programming logic often deviates from everyday common sense
- easy to give up when other more user-friendly tools are available
- difficult to express in code the idea of a solution to a problem



Why is it so difficult to start?

Learning a new language is never easy

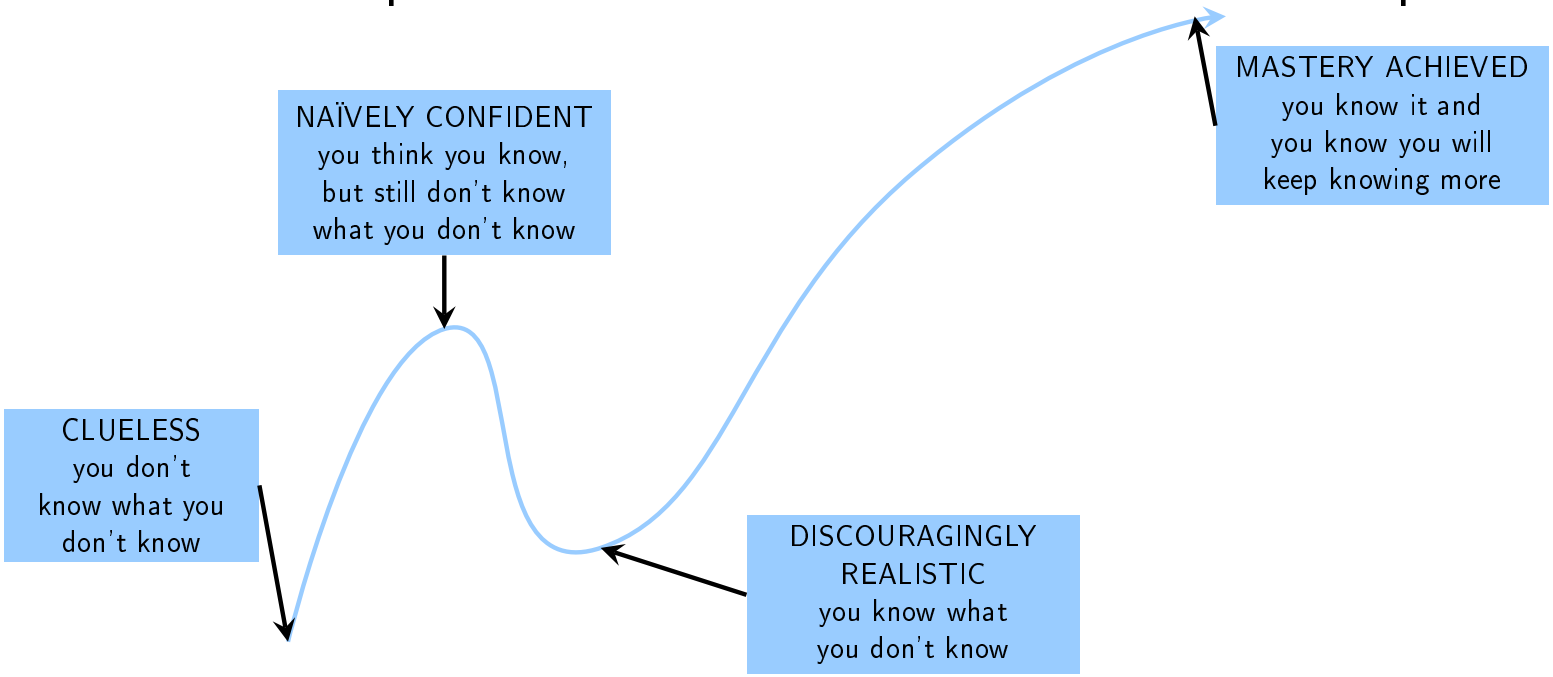
- develop some programming intuition
- master some key concepts of the language to communicate with the computer
- programming logic often deviates from everyday common sense
- easy to give up when other more user-friendly tools are available
- difficult to express in code the idea of a solution to a problem



Why is it so difficult to start?

Learning a new language is never easy

- develop some programming intuition
- master some key concepts of the language to communicate with the computer
- programming logic often deviates from everyday common sense
- easy to give up when other more user-friendly tools are available
- difficult to express in code the idea of a solution to a problem

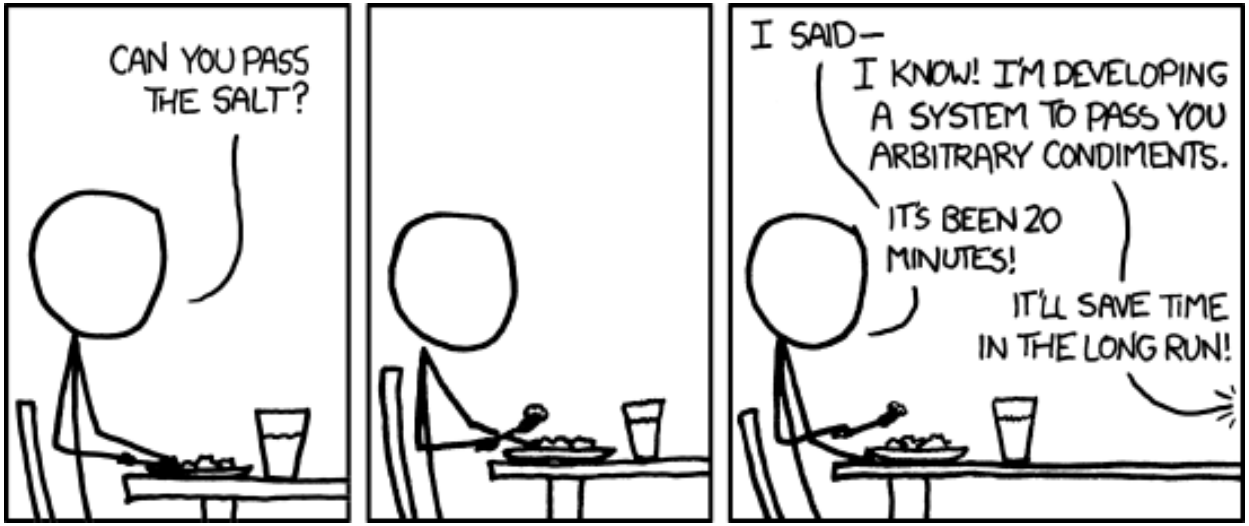


When should we program?

Programming may not always be the shortest path!

It depends on the predicted time to write the code and the complexity of the task to execute

Very difficult to find the threshold: depends mostly on the experience. If you have vast experience, you take less time to write code for trivial tasks or might already have coded in the past and re-use



What is R?

R is a language and environment for statistical computing and graphics

What is R?

R is a language and environment for statistical computing and graphics

- Programming language
- Compilation of methods to analyze data
- Graphical techniques

What is R?

R is a language and environment for statistical computing and graphics

- Programming language
- Compilation of methods to analyze data
- Graphical techniques

- You have to write (a lot of) code
- However, it is facilitated because someone else is doing the hard coding for you and sharing
- At the end you can plot your results as you wish (but you have to write more code)

What is R?

The Comprehensive R Archive Network (CRAN)

- R-Project (R programming language last version 3.5.1)
- R Packages repository (more than 10000 packages!)

What is R?

The Comprehensive R Archive Network (CRAN)

- R-Project (R programming language last version 3.5.1)
- R Packages repository (more than 10000 packages!)

Task views (http://cran.r-project.org/web/views/)

What is R?

The Comprehensive R Archive Network (CRAN)

- R-Project (R programming language last version 3.5.1)
- R Packages repository (more than 10000 packages!)

Task views (http://cran.r-project.org/web/views/)

Other packages repositories:

- Bioconductor http://www.bioconductor.org/
- R-forge - https://r-forge.r-project.org/ ATENTION!
- GitHub - https://github.com/trending?l=r ATENTION!

What is R?

POSITIVE

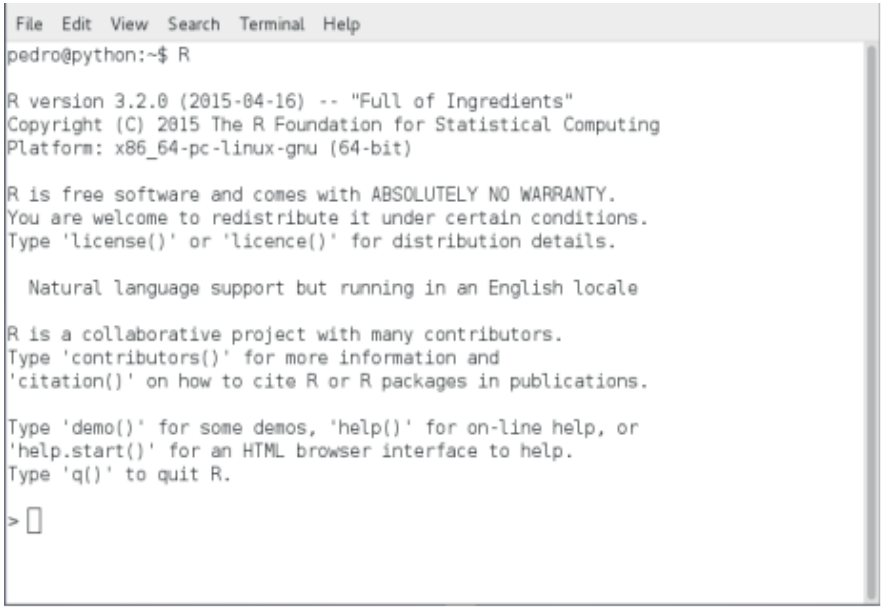
- Free and open source
- Becoming a standard for data analysis
- Clear syntax (easier to learn than other languages)
- Cross platform
- Modular with increasing number of packages
- Well documented and excellent on-line help with a vast community of users
- Methods to handle several sources of data (e.g. spatial, molecular, time-series)
- Several graphical interfaces available (general and dedicated)
- Scripting language (methods are written!)
- Excellent graphics
- Interpreted command line

NEGATIVE

- Steep learning curve...
- Several annoying idiosyncrasies (especially if you know other programming languages)
- Can be slow...
- Hard parallelization of the code
- Memory limitations

Install R?

1. Download R from CRAN (http://cran.r-project.org/) and install

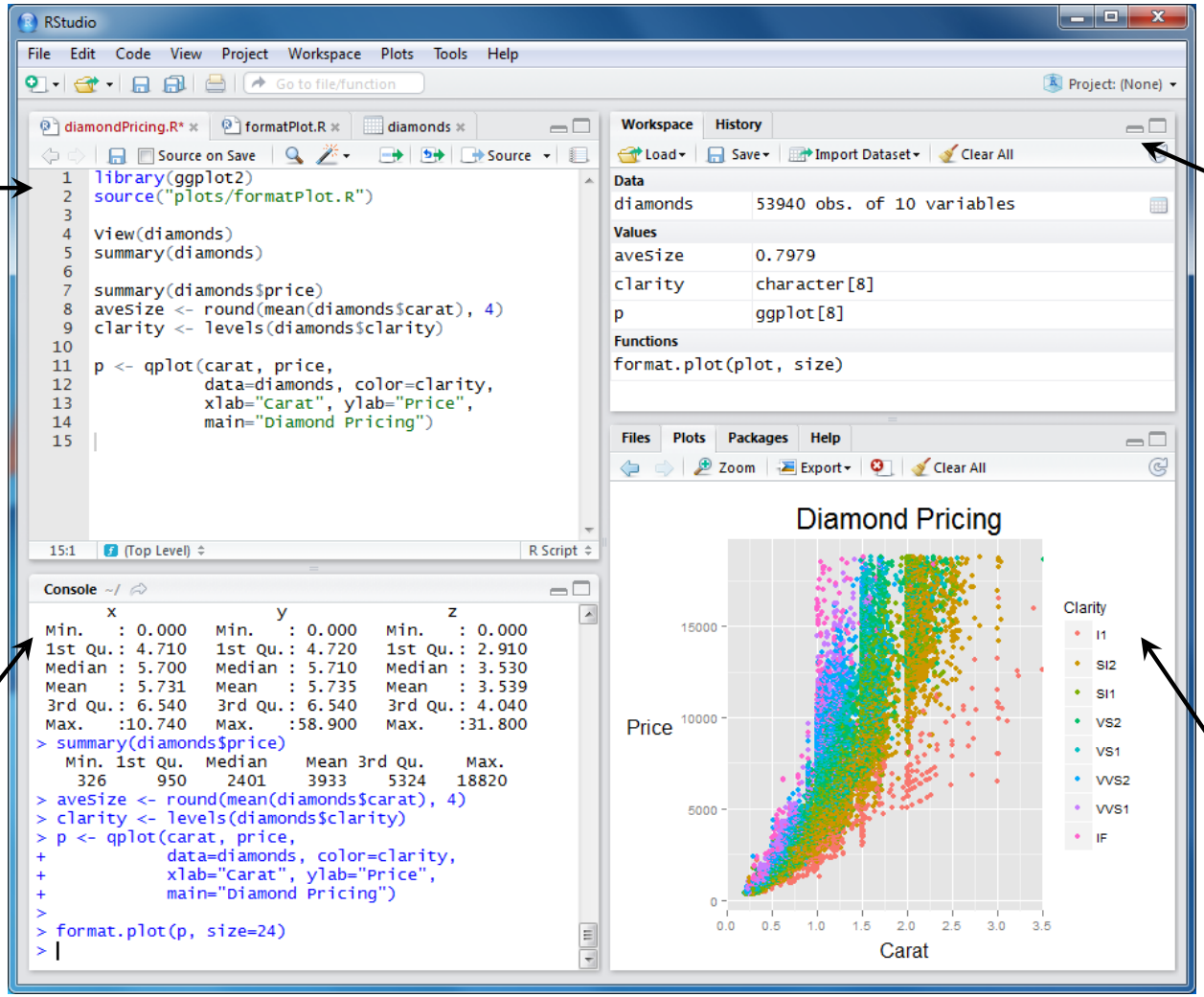


Suggestion

Free text editor: Notepad++ (http://notepad-plus-plus.org/)

Install R?

2. Download R-Studio (http://www.rstudio.com/products/rstudio/download/) and install



Editor
Object inspection

History
Objects in memory

Interactive
R session

Plotting
Help

Install R?

HELP

- CRAN Homepage - http://cran.r-project.org/
- R Seek - http://rseek.org/
- Mailing lists - http://www.r-project.org/mail.html
- Google – http://www.google.com

RESOURCES

- The R Journal - http://journal.r-project.org/
- R Bloggers - http://www.r-bloggers.com/
- Quick-R - http://www.statmethods.net/