# Software Archaeology

17-313 Fall 2024

Foundations of Software Engineering

https://cmu-313.github.io

Michael Hilton and Rohan Padhye

## Administrivia

- Make sure you're on Slack, and please add a profile picture.
    - See intro slides for other reminders (DM etiquette, office hours, etc).
- Homework 1 is released.
    - Part (a) is due Friday Aug 30th, 11:59 pm. That's tomorrow!
    - Part (b) is due Thursday, Sep 5th, 11:59pm.
    - This is an individual assignment; we will compose groups next week.
      **PLEASE FILL OUT TEAMWORK SURVEY**
    - Get started early, ask for help, and check the #technical-support channel; chances are your questions have been asked by others!

## Smoking Section

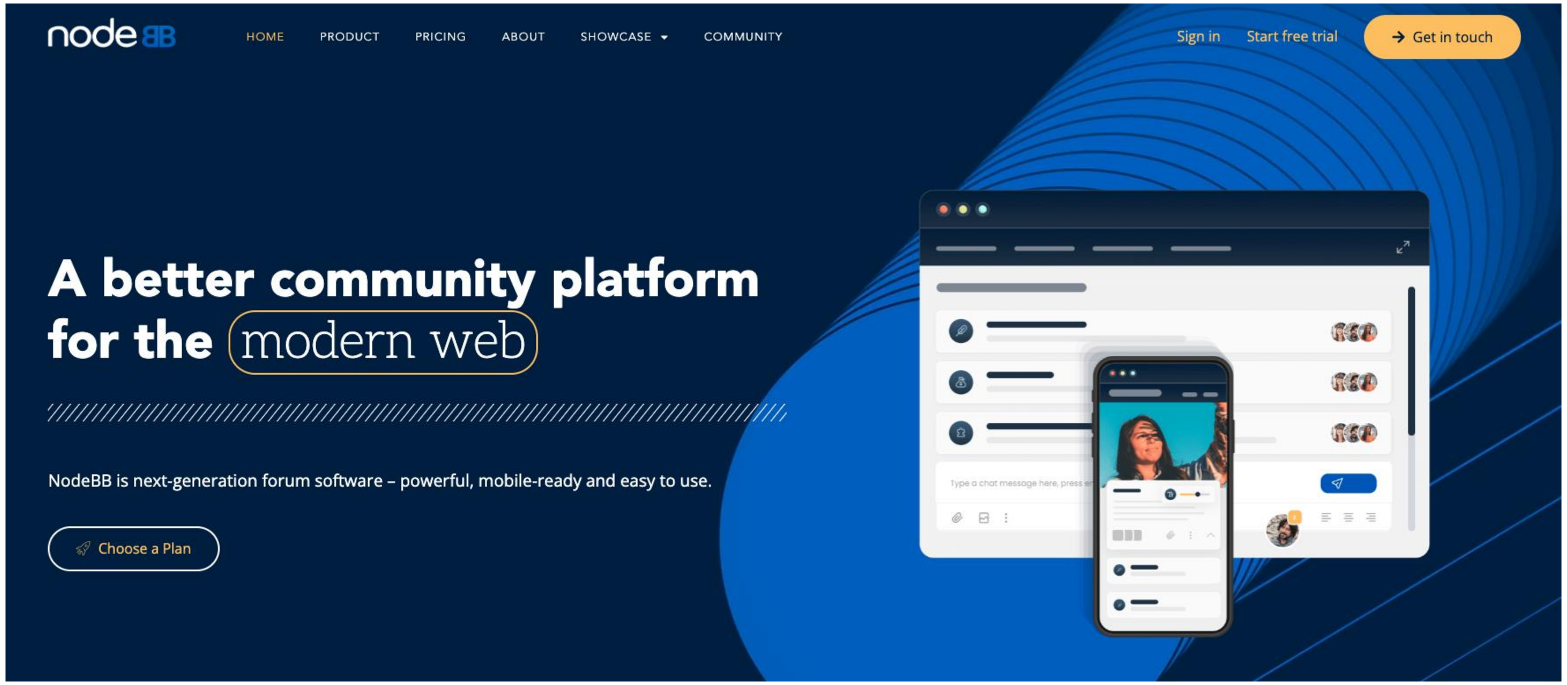
- Last full row

## Learning Goals

- Understand and scope the task of taking on and understanding a new and complex piece of existing software
- Appreciate the importance of configuring an effective IDE
- Contrast different types of code execution environments including local, remote, application, and libraries
- Enumerate both static and dynamic strategies for understanding and modifying a new codebase

# Context: big ole pile of code

- … do something with it!



A better community platform for the modern web

NodeBB is next-generation forum software – powerful, mobile-ready and easy to use.

## You will never understand the entire system!

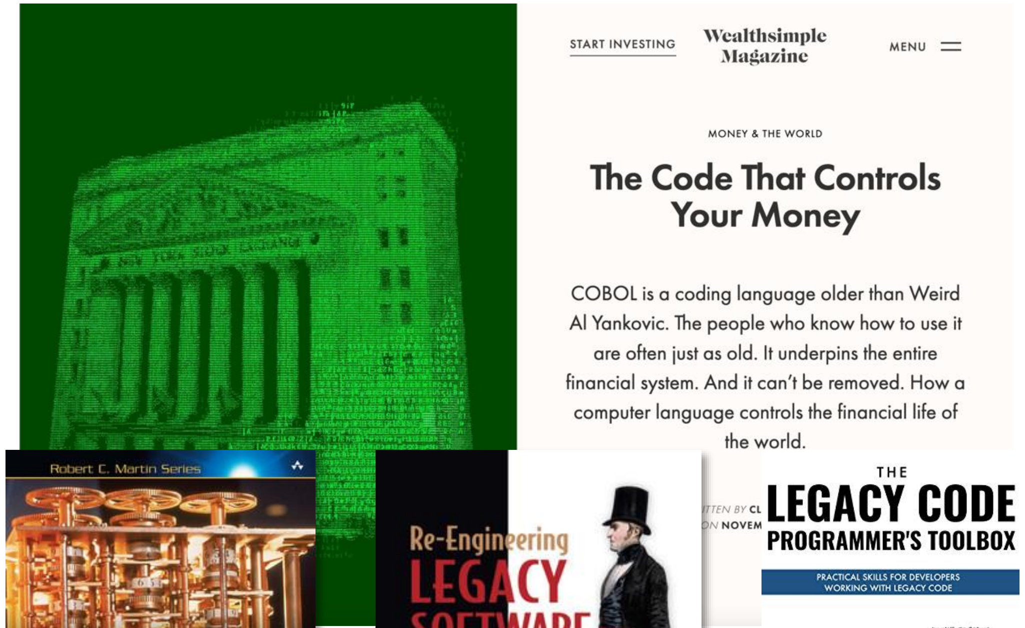
## Challenge: How do I tackle this codebase?

## Challenge: How do I tackle this codebase?

- Leverage your previous experiences (languages, technologies, patterns)
- Consult documentation, whitepapers
- Talk to experts, code owners
- Follow best practices to build a working model of the system

## Bad news: There are few helpful resources!

- **Working Effectively with Legacy Code.**
  Michael C. Feathers. 2004.
- **Re-Engineering Legacy Software.**
  Chris Birchall. 2016.
- **The Legacy Code Programmer's Toolbox.**
  Jonathan Boccara. 2019.



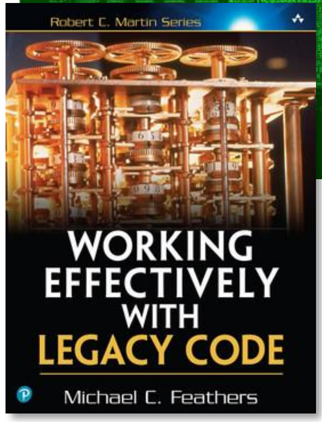
START INVESTING Wealthsimple Magazine MENU

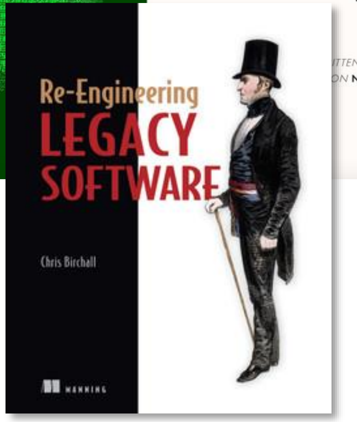
MONEY & THE WORLD

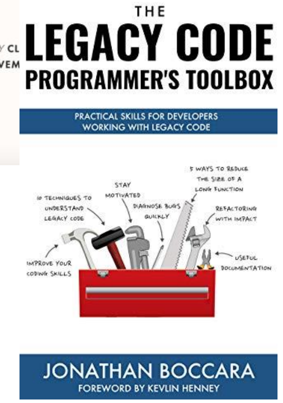
The Code That Controls Your Money

COBOL is a coding language older than Weird Al Yankovic. The people who know how to use it are often just as old. It underpins the entire financial system. And it can't be removed. How a computer language controls the financial life of the world.

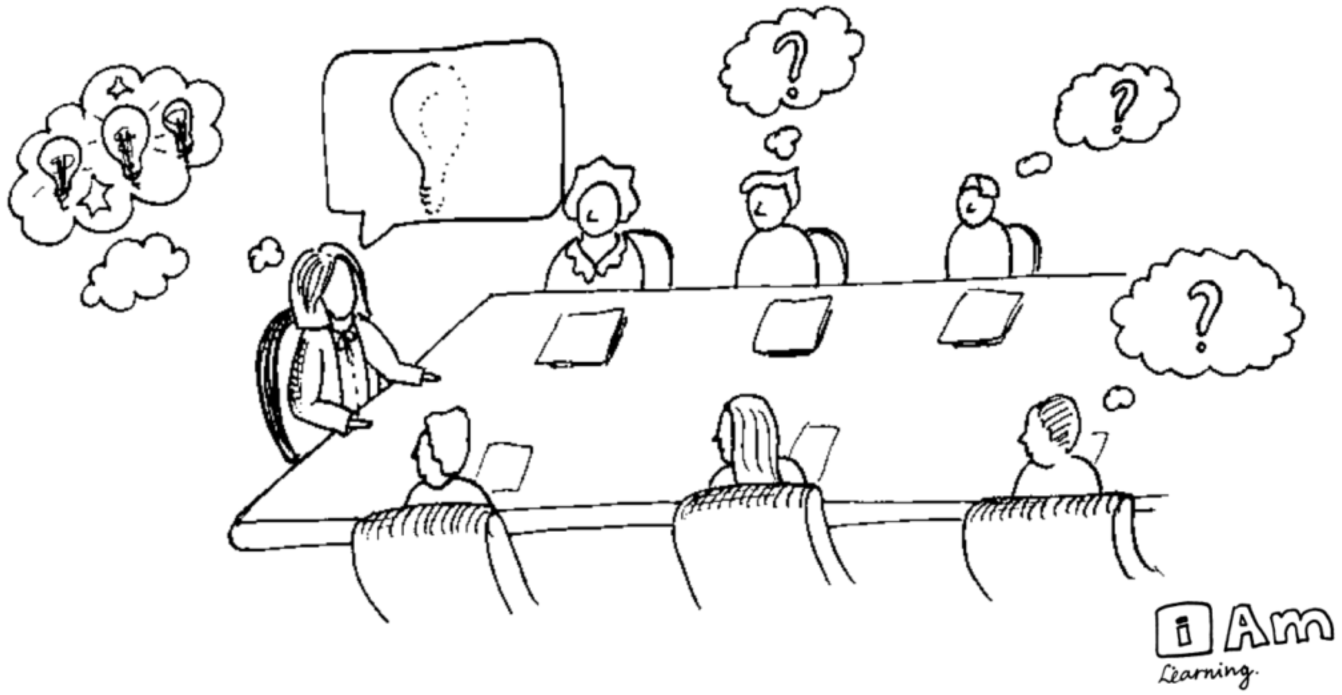
ITTEN BY CL
ON NOVEM





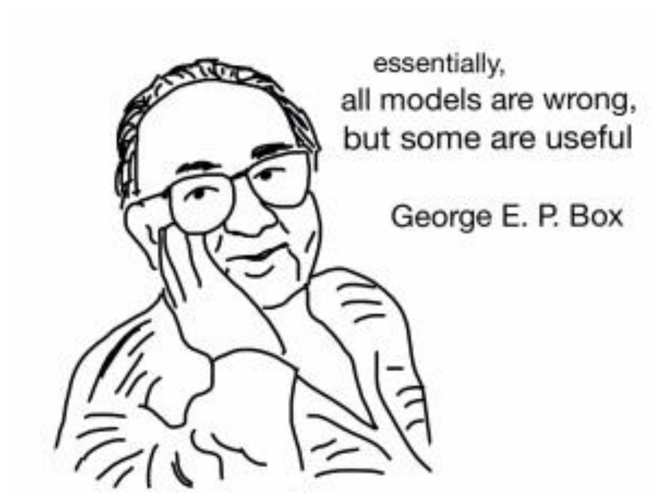

THE LEGACY CODE PROGRAMMER'S TOOLBOX
PRACTICAL SKILLS FOR DEVELOPERS WORKING WITH LEGACY CODE
JONATHAN BOCCARA
FOREWORD BY KEVLIN HENNEY

## Why? Because of Tacit Knowledge

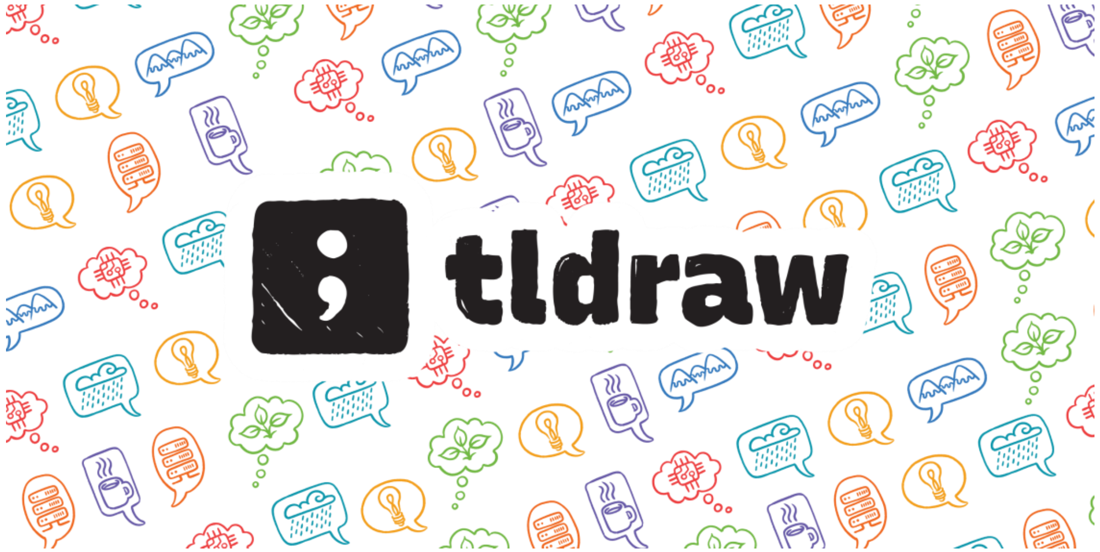


# Today: How to tackle codebases

- Goal: develop and test a working model or set of working hypotheses about how (some part of) a system works
- Working model: an understanding of the pieces of the system (components), and the way they interact (connections)
- Observation, probes, and hypothesis testing
  - Helpful tools and techniques!



## Live Demonstration: tldraw



[https://github.com/tldraw/tldraw](https://github.com/tldraw/tldraw)

## Steps to Understand a New Codebase

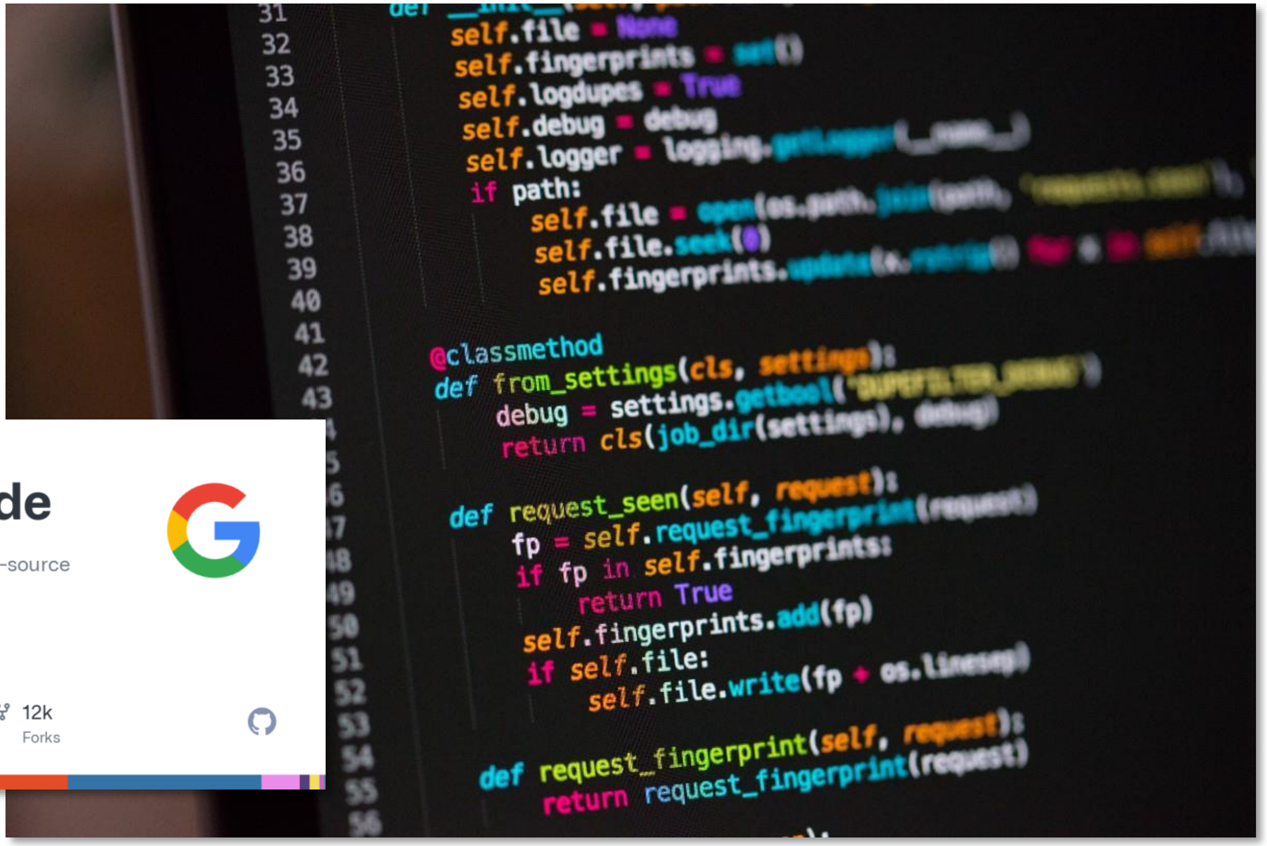
- Look at README.md
- Clone the repo.
- Build the codebase.
- Figure out how to make it run.
- What do you want to mess with?
    - Clone and own
- Traceability - Attach a debugger
    - View Source
    - Find the logs.
    - Search for constants (strings, colors, weird integers (#DEADBEEF))

# Participation Activity

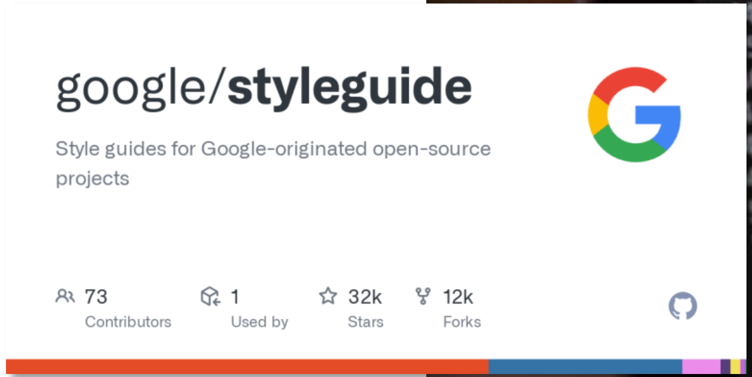
- Take out a piece of paper (or ask for one).
- Write down one pro and one con about trying to understand a new codebase by compiling and building it vs. just reading the code.
- Pair with your neighbor and discuss your answers. Do you agree?
- Share with the class!
- Write your own andrewID on the paper, leave it at the end of class.

# Observation: Software is full of patterns

- File structure
- System architecture
- Code structure
- Names
- …



```
def __init__(self, [illegible]
    self.file = None
    self.fingerprints = set()
    self.logdupes = True
    self.debug = debug
    self.logger = logging.getLogger(__name__)
    if path:
        self.file = open(os.path.join(path, 'requests.seen'), [illegible]
        self.file.seek(0)
        self.fingerprints.update([illegible]

@classmethod
def from_settings(cls, settings):
    debug = settings.getbool('DUPEFILTER_DEBUG')
    return cls(job_dir(settings), debug)

def request_seen(self, request):
    fp = self.request_fingerprint(request)
    if fp in self.fingerprints:
        return True
    self.fingerprints.add(fp)
    if self.file:
        self.file.write(fp + os.linesep)

def request_fingerprint(self, request):
    return request_fingerprint(request)
```



google/**styleguide**

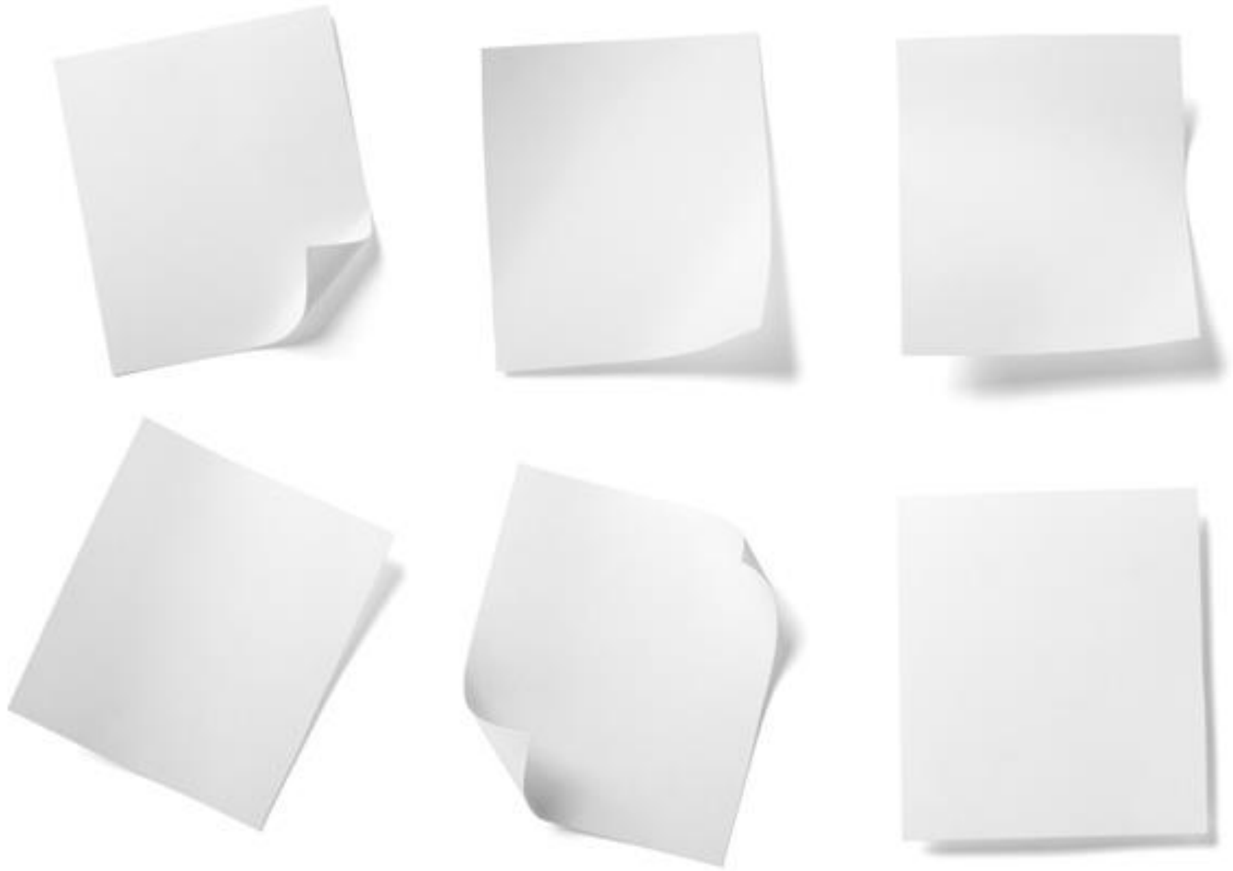
Style guides for Google-originated open-source projects

73 Contributors · 1 Used by · 32k Stars · 12k Forks

## Observation: Software is massively redundant

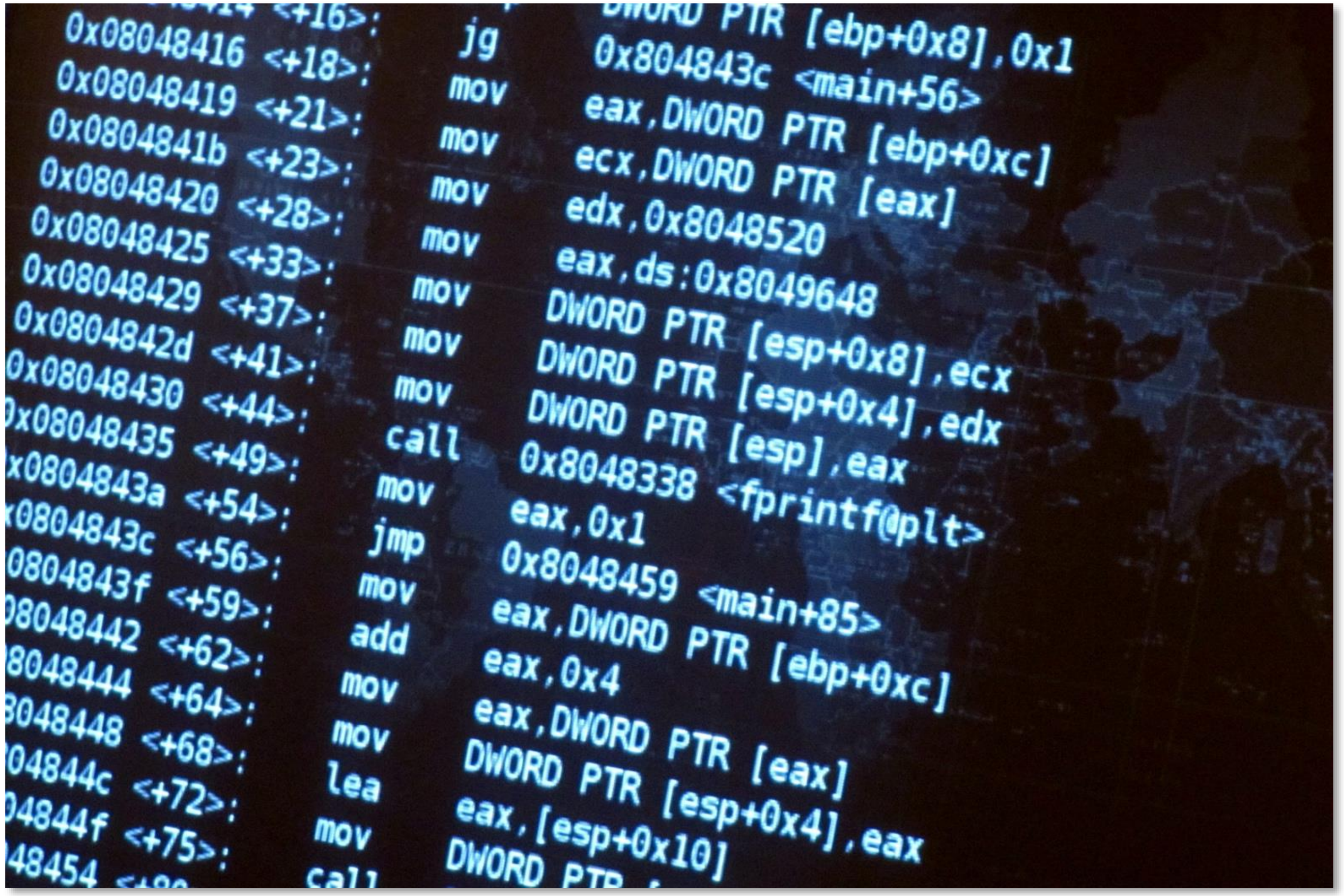
- There's always something to copy/use as a starting point!

## Observation: Code must run to do stuff!



## Observation: If code runs, it must have a beginning…

## Observation: If code runs, it must exist…



```
           <+16>:            DWORD PTR [ebp+0x8],0x1
0x08048416 <+18>:   jg       0x804843c <main+56>
0x08048419 <+21>:   mov      eax,DWORD PTR [ebp+0xc]
0x0804841b <+23>:   mov      ecx,DWORD PTR [eax]
0x08048420 <+28>:   mov      edx,0x8048520
0x08048425 <+33>:   mov      eax,ds:0x8049648
0x08048429 <+37>:   mov      DWORD PTR [esp+0x8],ecx
0x0804842d <+41>:   mov      DWORD PTR [esp+0x4],edx
0x08048430 <+44>:   mov      DWORD PTR [esp],eax
x08048435 <+49>:    call     0x8048338 <fprintf@plt>
x0804843a <+54>:    mov      eax,0x1
x0804843c <+56>:    jmp      0x8048459 <main+85>
0804843f <+59>:     mov      eax,DWORD PTR [ebp+0xc]
08048442 <+62>:     add      eax,0x4
8048444 <+64>:      mov      eax,DWORD PTR [eax]
8048448 <+68>:      mov      DWORD PTR [esp+0x4],eax
04844c <+72>:       lea      eax,[esp+0x10]
04844f <+75>:       mov      DWORD PTR
48454
```

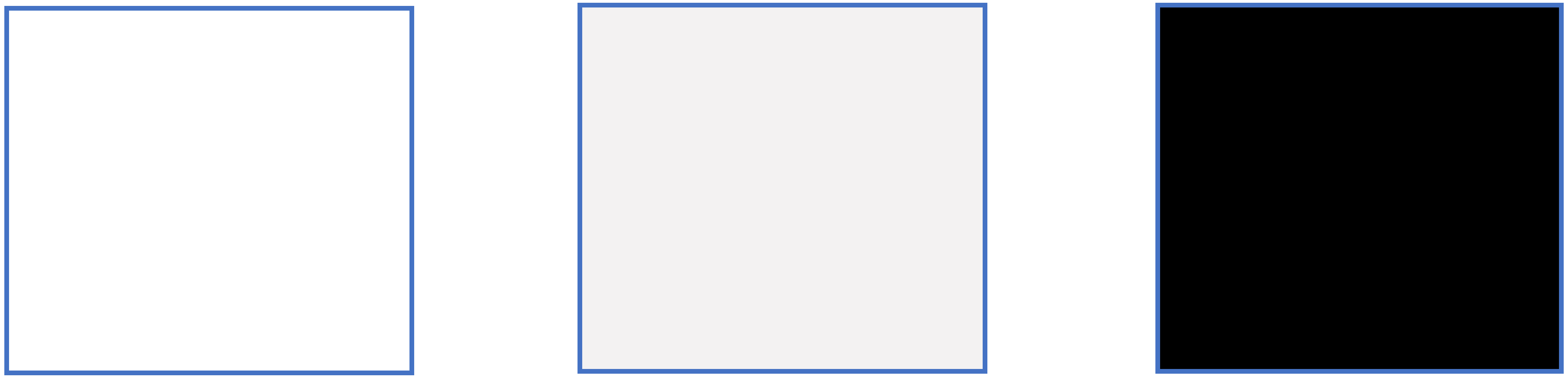
# The Beginning: Entry Points

- Locally installed programs: run cmd, OS launch, I/O events, etc.
- Local applications in dev: build + run, test, deploy (e.g., docker)
- Web apps server-side: Browser sends HTTP request (GET/POST)
- Web apps client-side: Browser runs JavaScript, event handlers

## Code must exist. But where?

- Locally installed programs: run cmd, OS launch, I/O events, etc.
  - Binaries (machine code) on your computer
- Local applications in dev: build + run, test, deploy (e.g., docker)
  - Source code in repository (+ dependencies)
- Web apps server-side: Browser sends HTTP request (e.g., GET, POST)
  - Code runs remotely (you can only observe outputs)
- Web apps client-side: Browser runs JavaScript, event handlers
  - Source code is downloaded and run locally (see: browser dev tools!)

# Can running code be Probed/Understood/Edited?



| White-box | Grey-box | | Black-box | |
|---|---|---|---|---|
| Source code built locally | Binaries running locally | | Server-side apps running remotely | |
| | Open source | Closed source | Open source | Closed source |
| (P+U+E) | (P+U) | (P) | (U) | (Talk to NSA) |

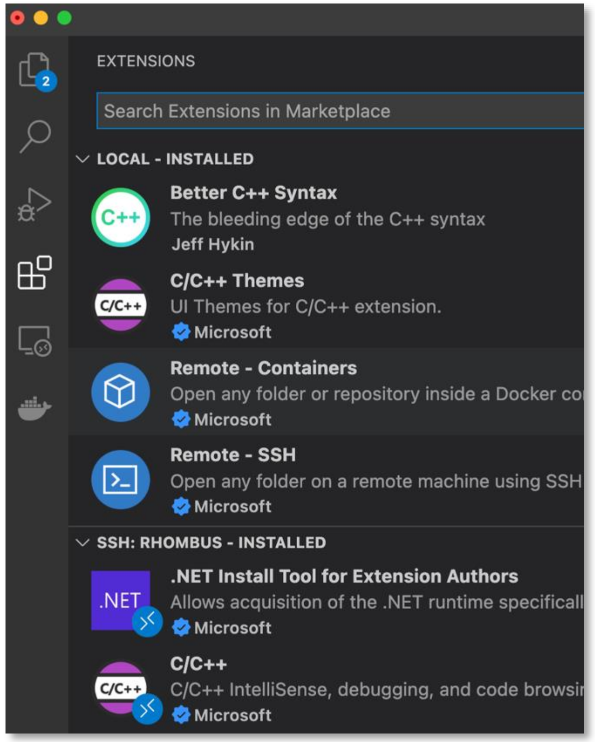
## Creating a model of unfamiliar code

Source code built locally

## Information Gathering

- Basic needs:
  - Code/file search and navigation
  - Code editing (probes)
  - Execution of code, tests
  - Observation of output (observation)
- Many choices here on tools! Depends on circumstance.
  - grep/find/etc.   Knowing Unix tools is invaluable
  - A decent IDE
  - Debugger
  - Test frameworks + coverage reports
  - Google (or your favorite web search engine)
  - ChatGPT or LaMA

# Static Information Gathering: Use an IDE!
# Real software is too complex to keep in your head



**Better C++ Syntax**
The bleeding edge of the C++ syntax
Jeff Hykin

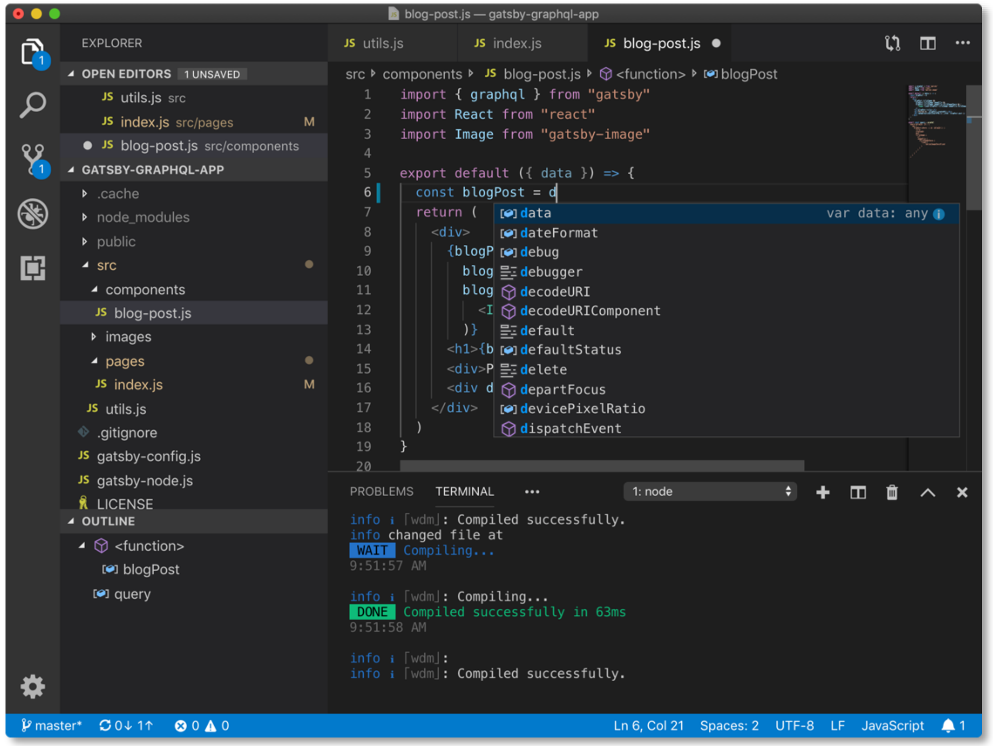
**C/C++ Themes**
UI Themes for C/C++ extension.
Microsoft

**Remote - Containers**
Open any folder or repository inside a Docker co
Microsoft

**Remote - SSH**
Open any folder on a remote machine using SSH
Microsoft

**.NET Install Tool for Extension Authors**
Allows acquisition of the .NET runtime specificall
Microsoft

**C/C++**
C/C++ IntelliSense, debugging, and code browsi
Microsoft



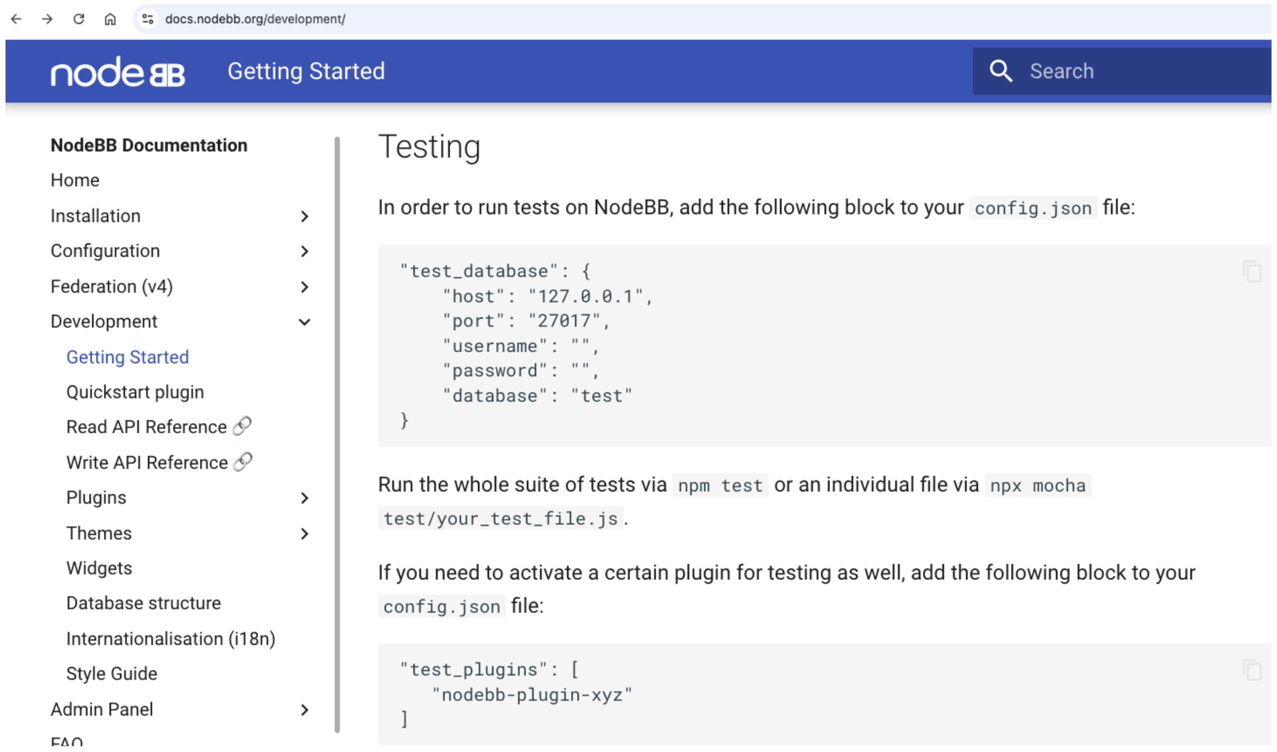
```
import { graphql } from "gatsby"
import React from "react"
import Image from "gatsby-image"

export default ({ data }) => {
  const blogPost = d
  return (   data                        var data: any
    <div>    dateFormat
      {blogP debug
        blog debugger
        blog decodeURI
          <I decodeURIComponent
        )}   default
      <h1>{b defaultStatus
      <div>P delete
      <div d departFocus
    </div>   devicePixelRatio
  )          dispatchEvent
}
```

```
info ℹ ｢wdm｣: Compiled successfully.
info changed file at
 WAIT  Compiling...
9:51:57 AM

info ℹ ｢wdm｣: Compiling...
 DONE  Compiled successfully in 63ms
9:51:58 AM

info ℹ ｢wdm｣:
info ℹ ｢wdm｣: Compiled successfully.
```

master* 0↓ 1↑ ⊗ 0 ⚠ 0 Ln 6, Col 21 Spaces: 2 UTF-8 LF JavaScript 🔔 1

# Consider documentation and tutorials judiciously

- Great for discovering entry points!
- Can teach you about general structure, architecture (more on this later in the semester)
- Often out of date.
- As you gain experience, you will recognize more of these, and you will immediately know something about how the program works
- Also: discussion boards; issue trackers



Testing

In order to run tests on NodeBB, add the following block to your `config.json` file:

```
"test_database": {
    "host": "127.0.0.1",
    "port": "27017",
    "username": "",
    "password": "",
    "database": "test"
}
```

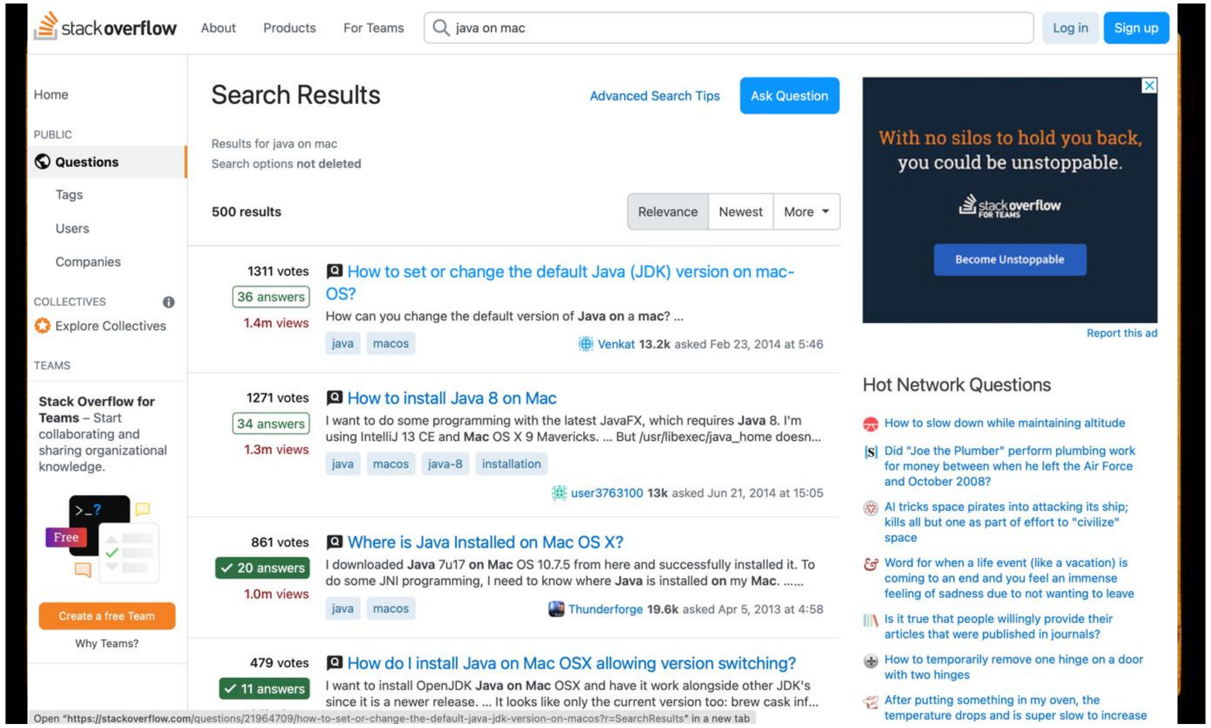
Run the whole suite of tests via `npm test` or an individual file via `npx mocha test/your_test_file.js`.

If you need to activate a certain plugin for testing as well, add the following block to your `config.json` file:

```
"test_plugins": [
    "nodebb-plugin-xyz"
]
```

# Discussion Boards and Issue Trackers

- Software is written by people.
- How can we talk to them?
- Fortunately, they probably aren't dead.
- So, you can report problems on GitHub.
- Or, ask them questions on StackOverflow.



stackoverflow About Products For Teams java on mac Log in Sign up

Home

PUBLIC

Questions

Tags

Users

Companies

COLLECTIVES

Explore Collectives

TEAMS

**Stack Overflow for Teams** – Start collaborating and sharing organizational knowledge.

Free

Create a free Team

Why Teams?

Search Results

Advanced Search Tips Ask Question

Results for java on mac

Search options **not deleted**

**500 results**

Relevance Newest More

1311 votes 36 answers 1.4m views
How to set or change the default Java (JDK) version on macOS?
How can you change the default version of Java on a mac? ...
java macos
Venkat **13.2k** asked Feb 23, 2014 at 5:46

1271 votes 34 answers 1.3m views
How to install Java 8 on Mac
I want to do some programming with the latest JavaFX, which requires **Java** 8. I'm using IntelliJ 13 CE and **Mac** OS X 9 Mavericks. ... But /usr/libexec/java_home doesn...
java macos java-8 installation
user3763100 **13k** asked Jun 21, 2014 at 15:05

861 votes ✓ 20 answers 1.0m views
Where is Java Installed on Mac OS X?
I downloaded **Java** 7u17 **on Mac** OS 10.7.5 from here and successfully installed it. To do some JNI programming, I need to know where **Java** is installed **on** my **Mac**. ......
java macos
Thunderforge **19.6k** asked Apr 5, 2013 at 4:58

479 votes ✓ 11 answers
How do I install Java on Mac OSX allowing version switching?
I want to install OpenJDK **Java on Mac** OSX and have it work alongside other JDK's since it is a newer release. ... It looks like only the current version too: brew cask inf...

With no silos to hold you back, you could be unstoppable.

stackoverflow FOR TEAMS

Become Unstoppable

Report this ad

Hot Network Questions

How to slow down while maintaining altitude

Did "Joe the Plumber" perform plumbing work for money between when he left the Air Force and October 2008?

AI tricks space pirates into attacking its ship; kills all but one as part of effort to "civilize" space

Word for when a life event (like a vacation) is coming to an end and you feel an immense feeling of sadness due to not wanting to leave

Is it true that people willingly provide their articles that were published in journals?

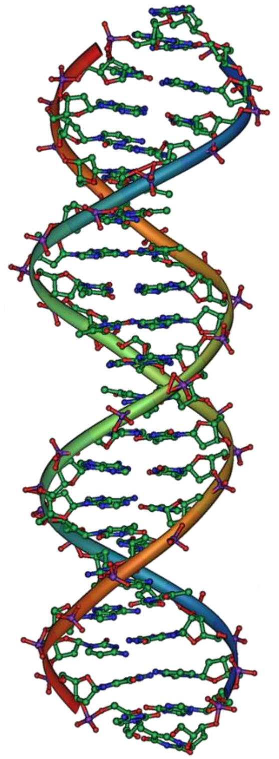
How to temporarily remove one hinge on a door with two hinges

After putting something in my oven, the temperature drops and is super slow to increase

Open "https://stackoverflow.com/questions/21964709/how-to-set-or-change-the-default-java-jdk-version-on-macos?r=SearchResults" in a new tab

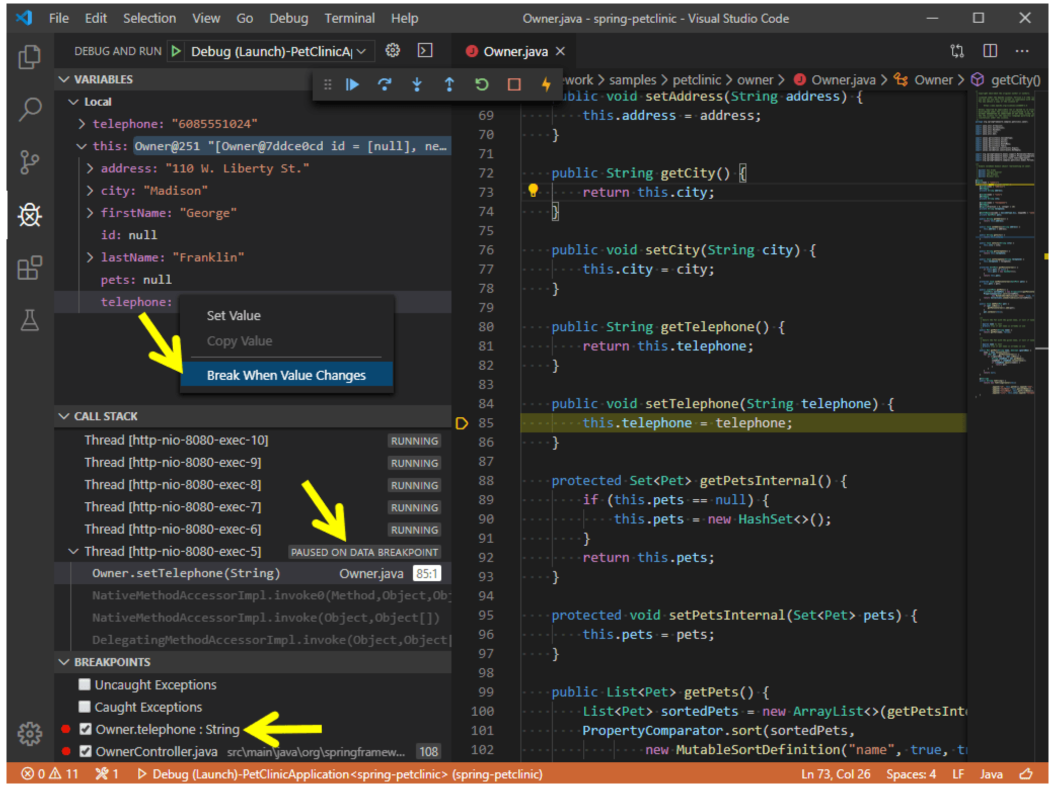
# Dynamic Information Gathering
Change helps to inform and refine mental models

- Build it.
- Run it.
- Change it.
- Run it again.
- How did the behavior change?



# Probes: Observe, control or “lightly” manipulate execution

- print(“this code is running!”)
- Structured logging
- Debuggers
  - Breakpoint, eval, step through / step over
  - (Some tools even support remote debugging)
- Delete debugging
- Chrome Developer Tools


File Edit Selection View Go Debug Terminal Help
Owner.java - spring-petclinic - Visual Studio Code
DEBUG AND RUN
Debug (Launch)-PetClinicA|
Owner.java
work > samples > petclinic > owner > Owner.java > Owner > getCity()
Set Value
Copy Value
Break When Value Changes

```
VARIABLES
  Local
    > telephone: "6085551024"
    v this: Owner@251 "[Owner@7ddce0cd id = [null], ne…
      > address: "110 W. Liberty St."
      > city: "Madison"
      > firstName: "George"
        id: null
      > lastName: "Franklin"
        pets: null
        telephone:

CALL STACK
  Thread [http-nio-8080-exec-10]        RUNNING
  Thread [http-nio-8080-exec-9]         RUNNING
  Thread [http-nio-8080-exec-8]         RUNNING
  Thread [http-nio-8080-exec-7]         RUNNING
  Thread [http-nio-8080-exec-6]         RUNNING
v Thread [http-nio-8080-exec-5]   PAUSED ON DATA BREAKPOINT
    Owner.setTelephone(String)          Owner.java 85:1
    NativeMethodAccessorImpl.invoke0(Method,Object,Ob
    NativeMethodAccessorImpl.invoke(Object,Object[])
    DelegatingMethodAccessorImpl.invoke(Object,Object|

BREAKPOINTS
  ☐ Uncaught Exceptions
  ☐ Caught Exceptions
  ● ☑ Owner.telephone : String
  ● ☑ OwnerController.java  src\main\java\org\springframew...  108
```

```java
    ublic void setAddress(String address) {
        this.address = address;
    }

    public String getCity() {
        return this.city;
    }

    public void setCity(String city) {
        this.city = city;
    }

    public String getTelephone() {
        return this.telephone;
    }

    public void setTelephone(String telephone) {
        this.telephone = telephone;
    }

    protected Set<Pet> getPetsInternal() {
        if (this.pets == null) {
            this.pets = new HashSet<>();
        }
        return this.pets;
    }

    protected void setPetsInternal(Set<Pet> pets) {
        this.pets = pets;
    }

    public List<Pet> getPets() {
        List<Pet> sortedPets = new ArrayList<>(getPetsInt
        PropertyComparator.sort(sortedPets,
                new MutableSortDefinition("name", true, t
```

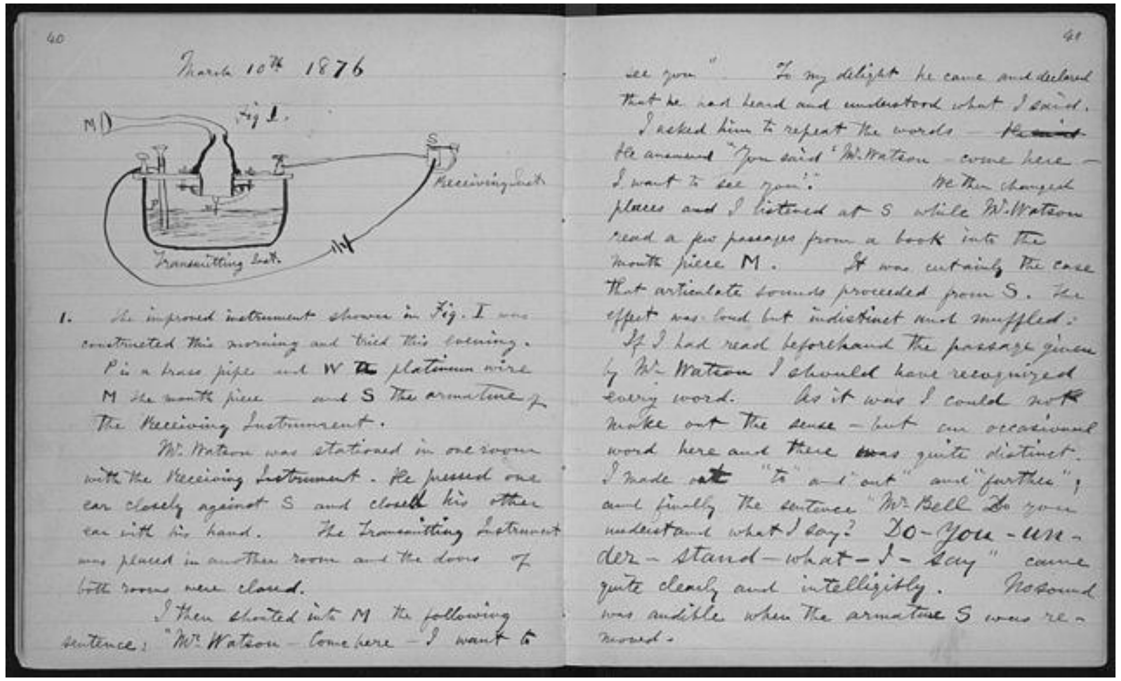
```
⊗ 0 ⚠ 11   1   Debug (Launch)-PetClinicApplication<spring-petclinic> (spring-petclinic)   Ln 73, Col 26   Spaces: 4   LF   Java
```

## Step 0: Sanity check basic model + hypotheses

- Confirm that you can build and run the code.
  - Ideally both using the tests provided, and by hand.
- Confirm that the code you are running is the code you built
- Confirm that you can make an externally visible change
- How? Where? Starting points:
  - Run an existing test, change it
  - Write a new test
  - Change the code, write or rerun a test that should notice the change
- Ask someone for help

# Document and share your findings!

- Update README and docs
  - Or better: use a Developer Wiki
  - Use Mermaid for diagrams
- Screencast on Twitch
- Collaborate with others
- Include negative results, too!



40

March 10th 1876

Fig 1.

M
S
P
W
Receiving Inst.
Transmitting Inst.

1. The improved instrument shown in Fig. I was constructed this morning and tried this evening. P is a brass pipe and W ~~the~~ platinum wire M the mouth piece and S the armature of the Receiving Instrument.

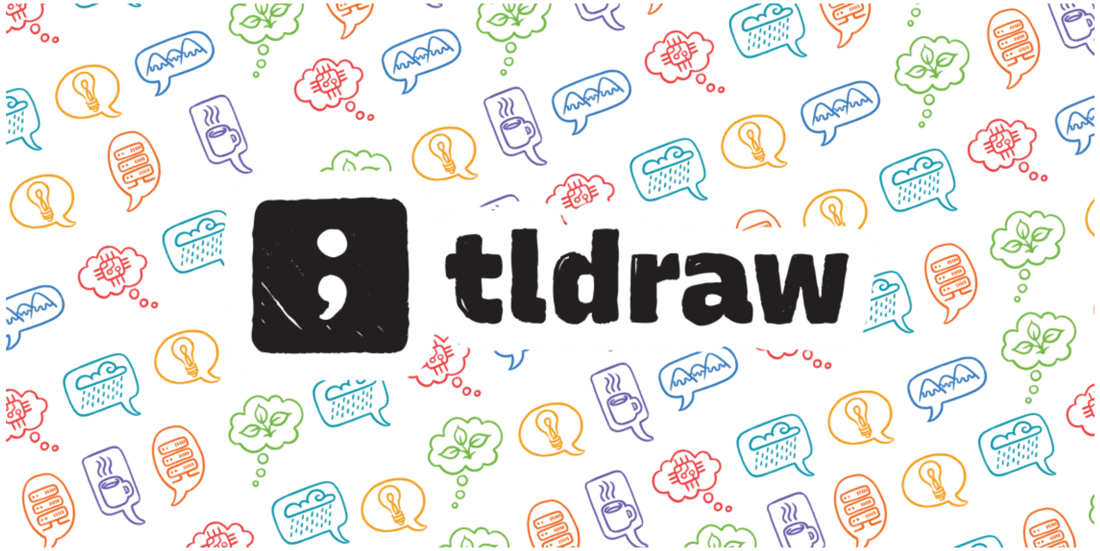
Mr. Watson was stationed in one room with the Receiving Instrument. He pressed one ear closely against S and closed his other ear with his hand. The Transmitting Instrument was placed in another room and the doors of both rooms were closed.

I then shouted into M the following sentence: "Mr. Watson – Come here – I want to

41

see you." To my delight he came and declared that he had heard and understood what I said. I asked him to repeat the words – ~~He said~~ He answered "You said 'Mr. Watson – come here – I want to see you.'" We then changed places and I listened at S while Mr. Watson read a few passages from a book into the mouth piece M. It was certainly the case that articulate sounds proceeded from S. The effect was loud but indistinct and muffled. If I had read beforehand the passage given by Mr. Watson I should have recognized every word. As it was I could not make out the sense – but an occasional word here and there ~~was~~ quite distinct. I made ~~out~~ "to" and "out" and "further"; and finally the sentence "Mr. Bell Do you understand what I say? Do – you – un – der – stand – what – I – say" came quite clearly and intelligibly. No sound was audible when the armature S was removed.

## Let’s try some of these techniques again…



https://github.com/tldraw/tldraw

## Next time…

- Boeing 737 MAX Case Study
- Do the assigned reading before class! Small Canvas quiz.