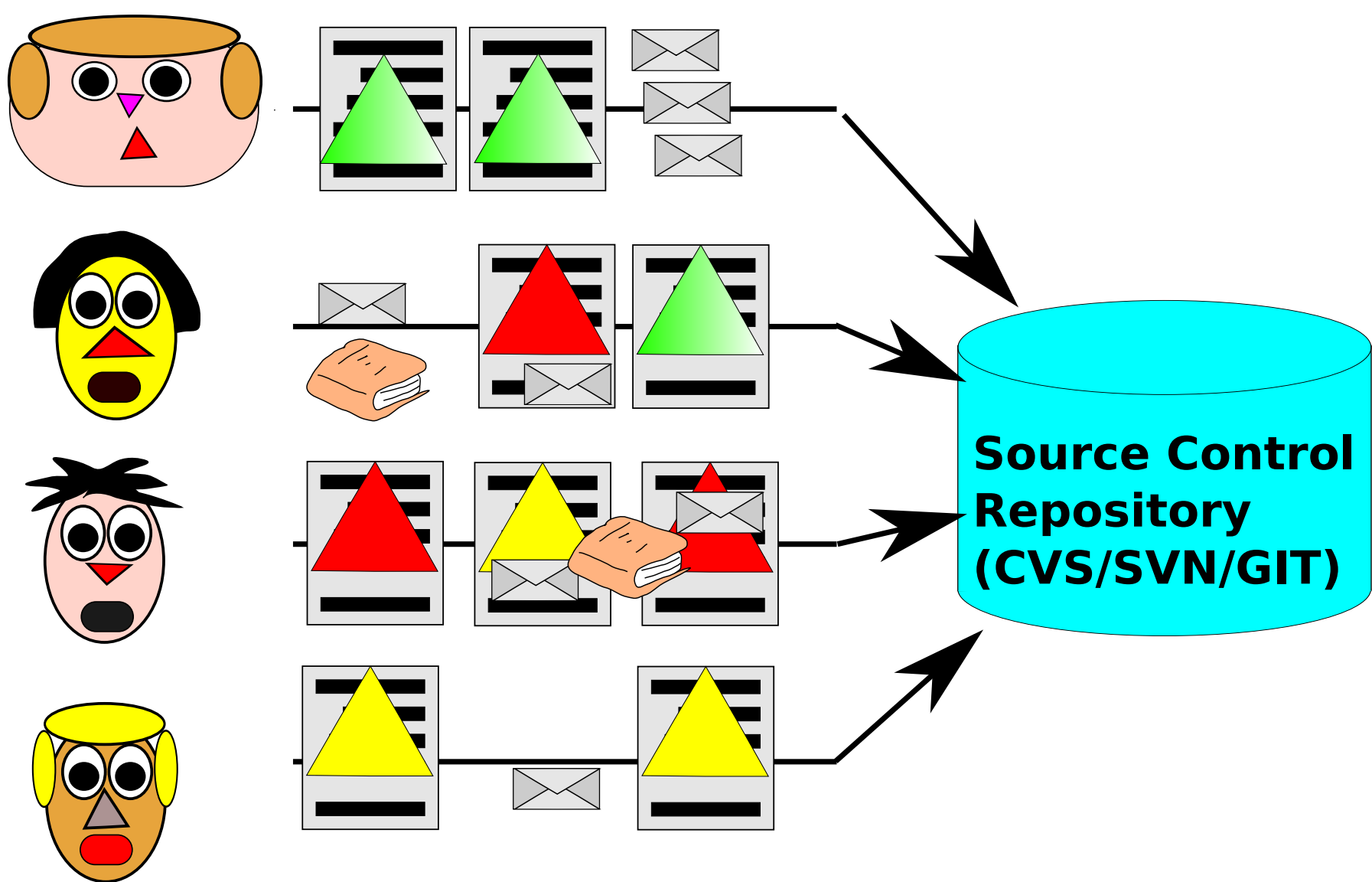


Mining Recurrent Activities: Fourier Analysis of Change Events

Abram Hindle, Michael Godfrey, Ric Holt
University Of Waterloo, Waterloo, Canada

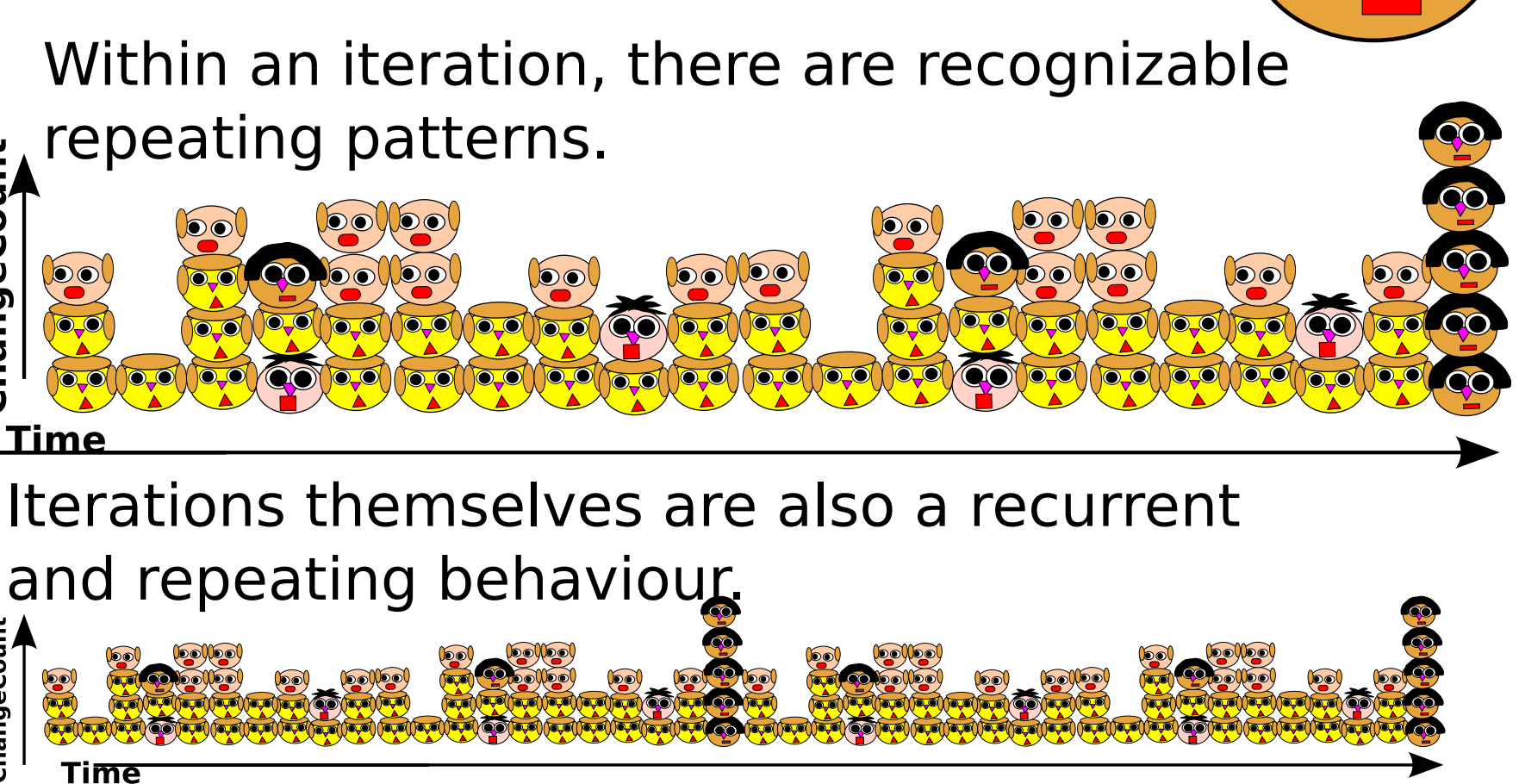
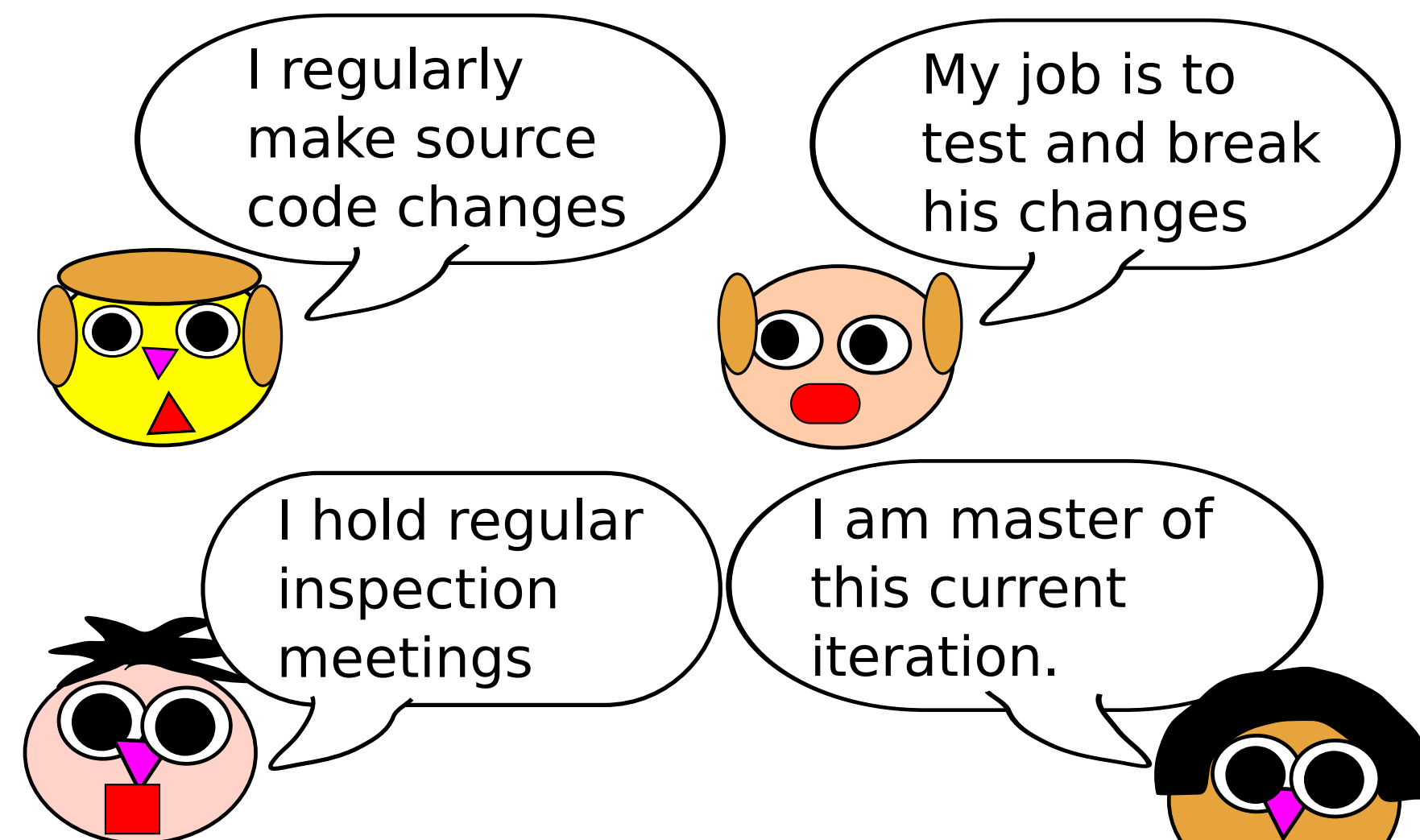
Developers create artifacts



Developers create various kinds of artifacts when they make changes:

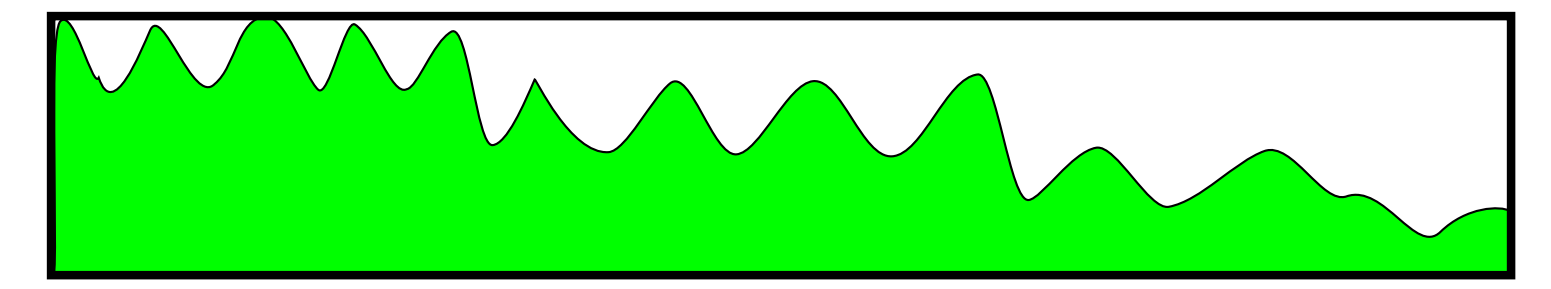
source code, test suites, bug reports, documentation, mailing list messages, etc

Software development has recurrent behaviour



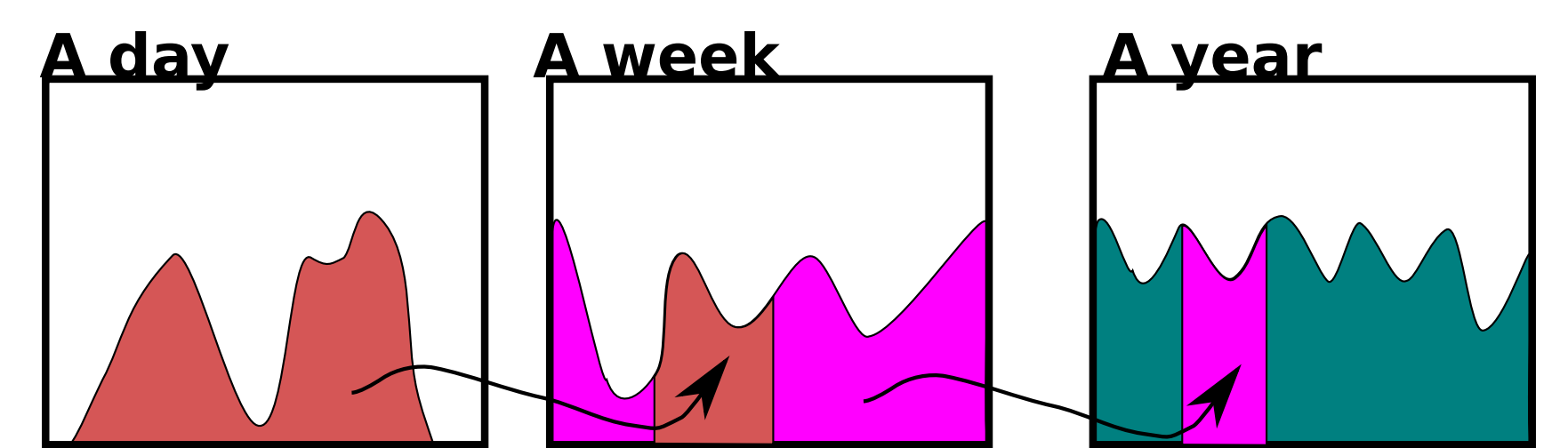
How do we discover recurrent behaviour? With Time-series analysis

A signal, multiple years in length



You must choose a period to use time-series analysis!

ASSUME A PERIOD

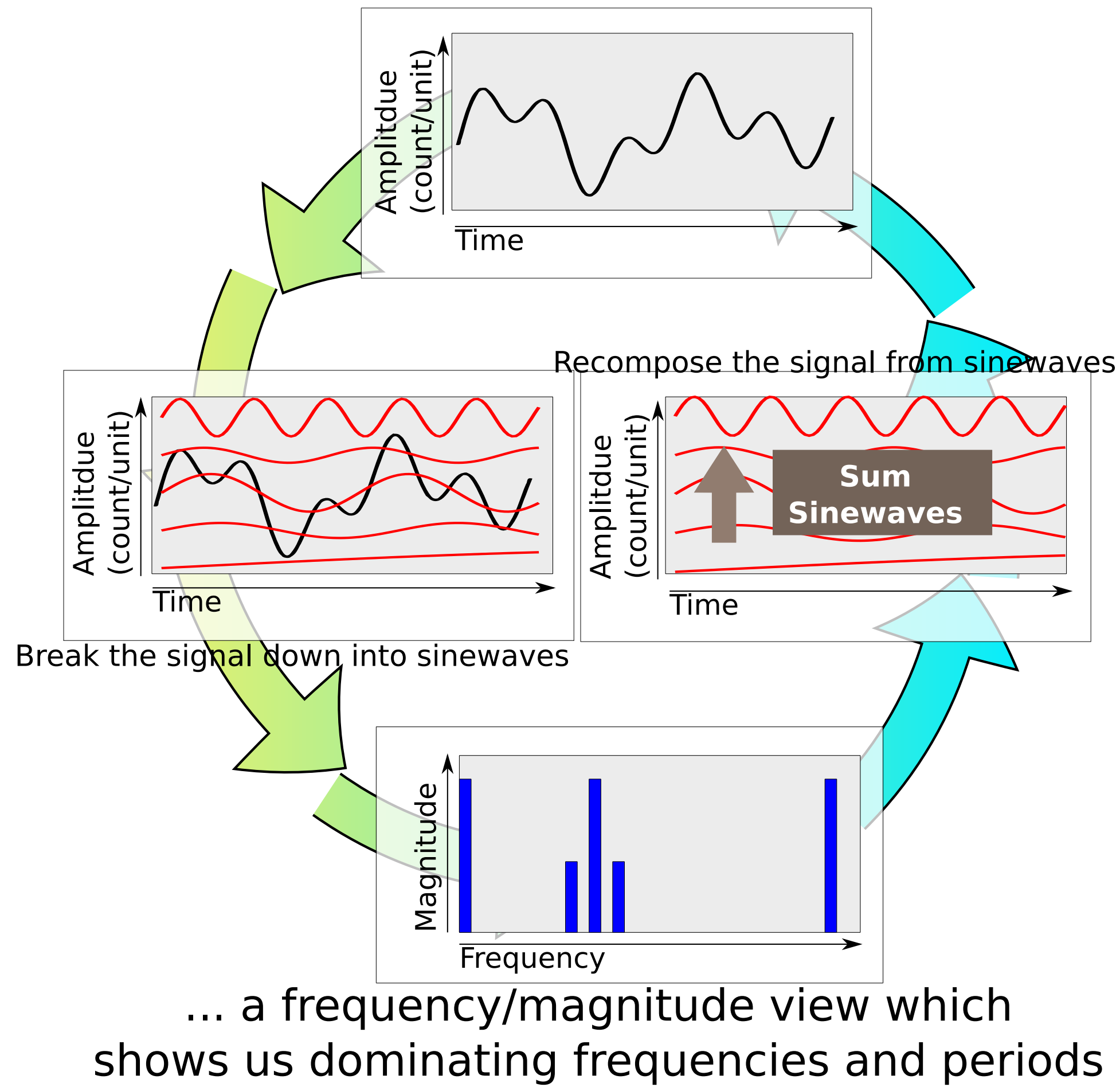


But which is correct or useful?

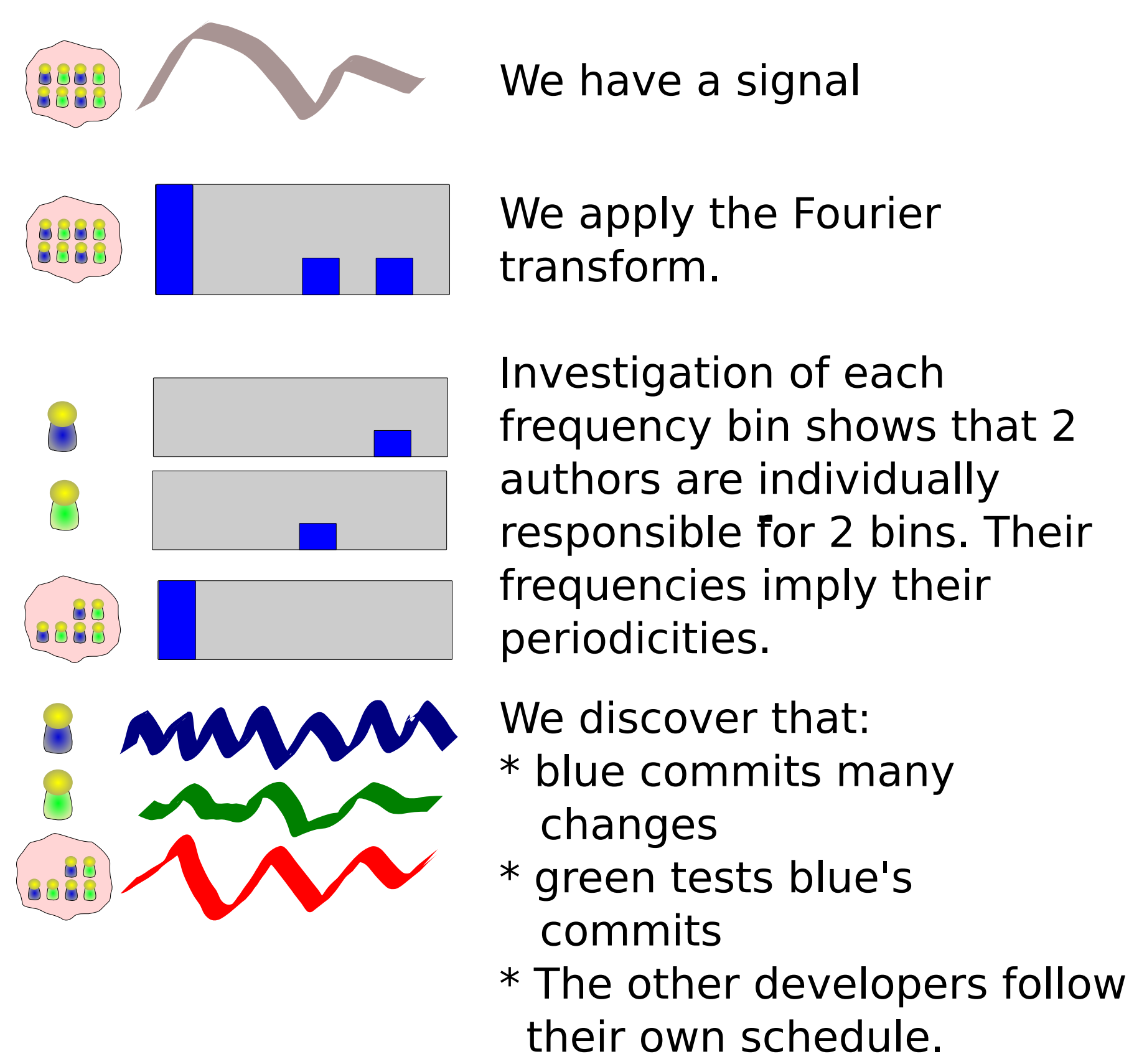
What period should I analyze this signal by? If only I had a tool to tell me what do!

A solution: Fourier transform

We can convert from a time/amplitude view to...

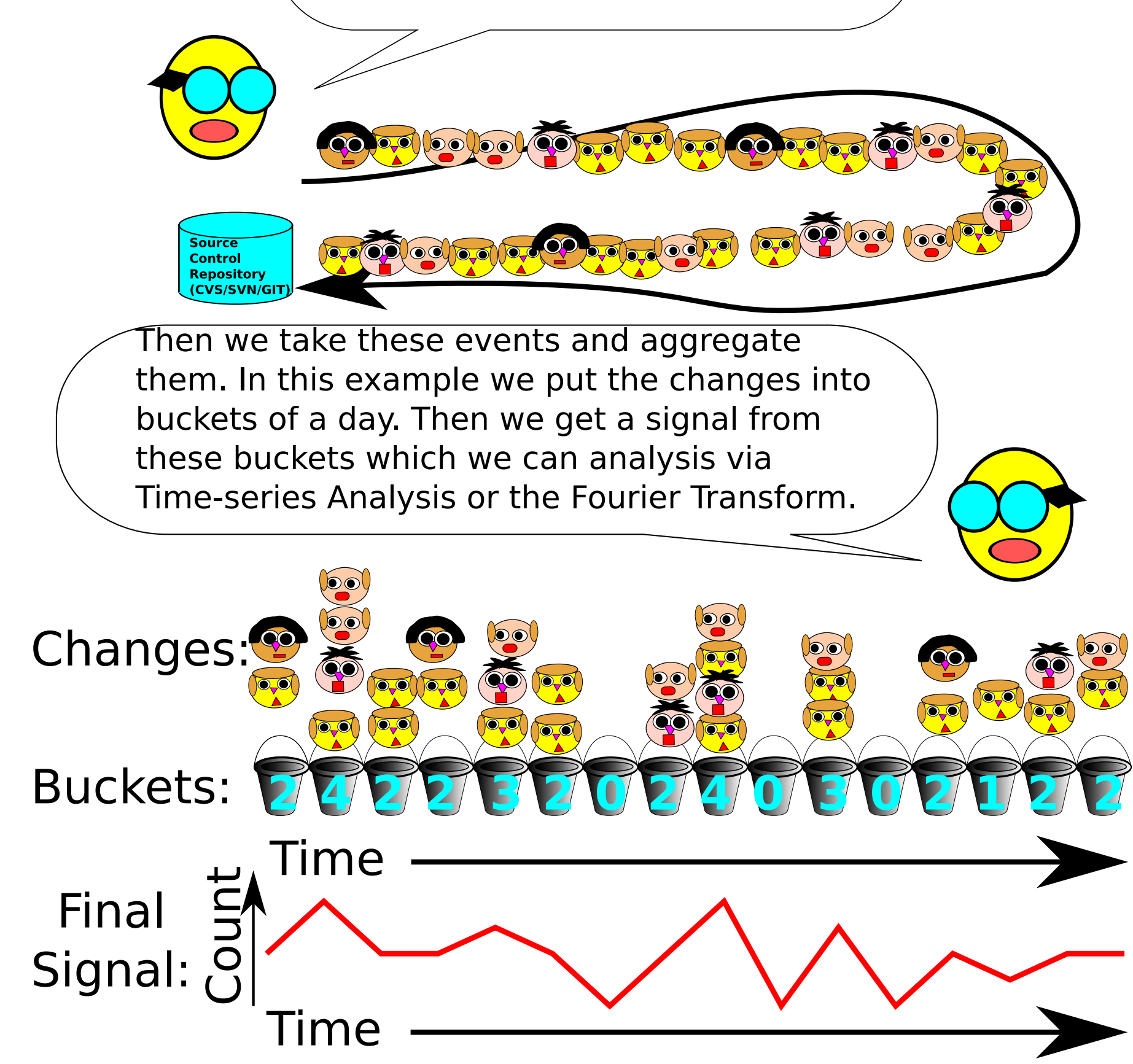


How can we apply the Fourier transform to software change and software related data?



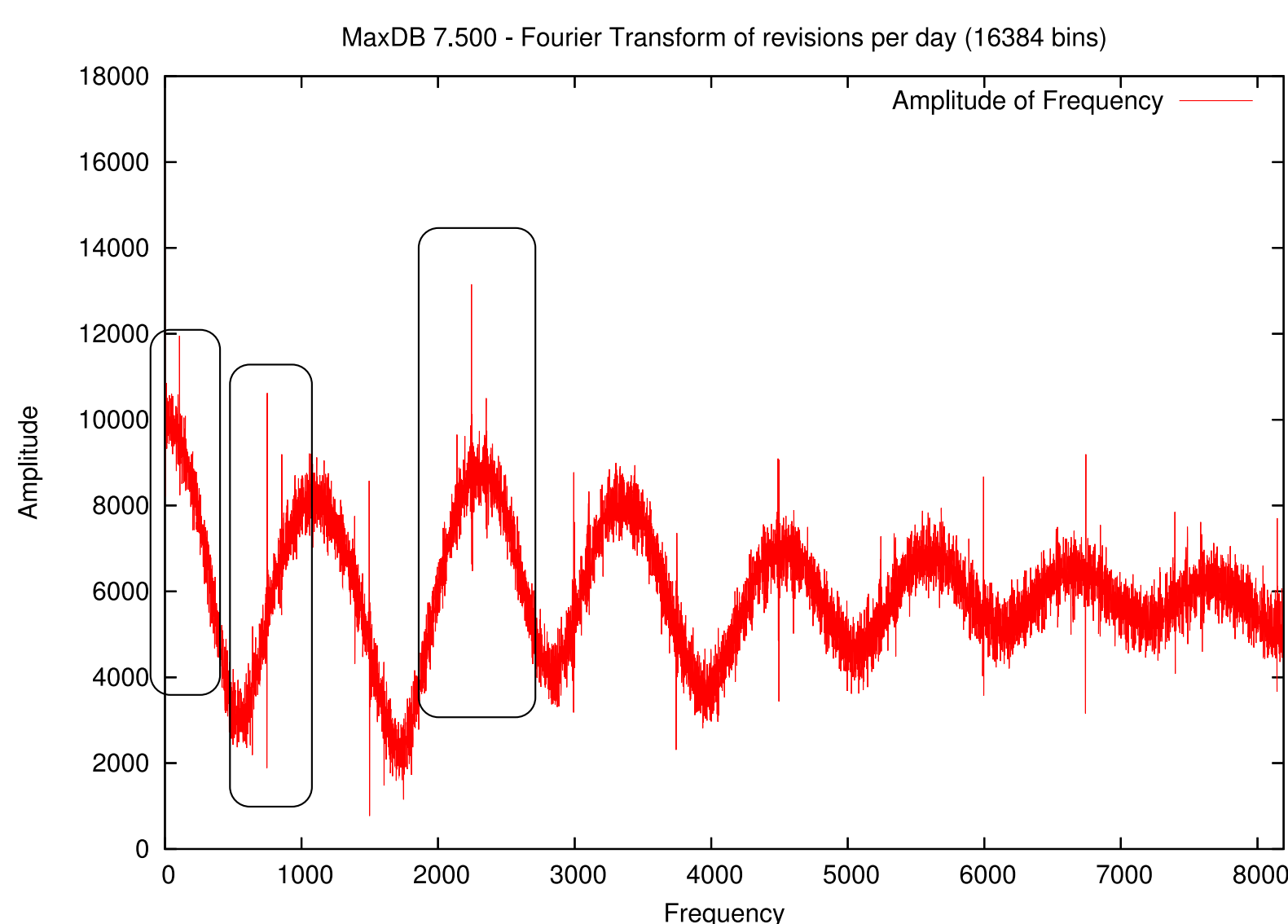
Convert discrete events into signals

First, take an event stream, such as, revisions to a project, log events, mailing list posts, etc.



So what? Does it work?

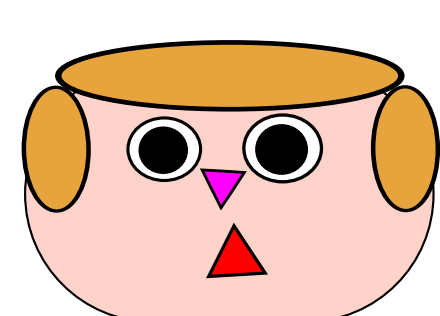
We can find periodicities of software. Here's the Fourier transform of MaxDB 7.500



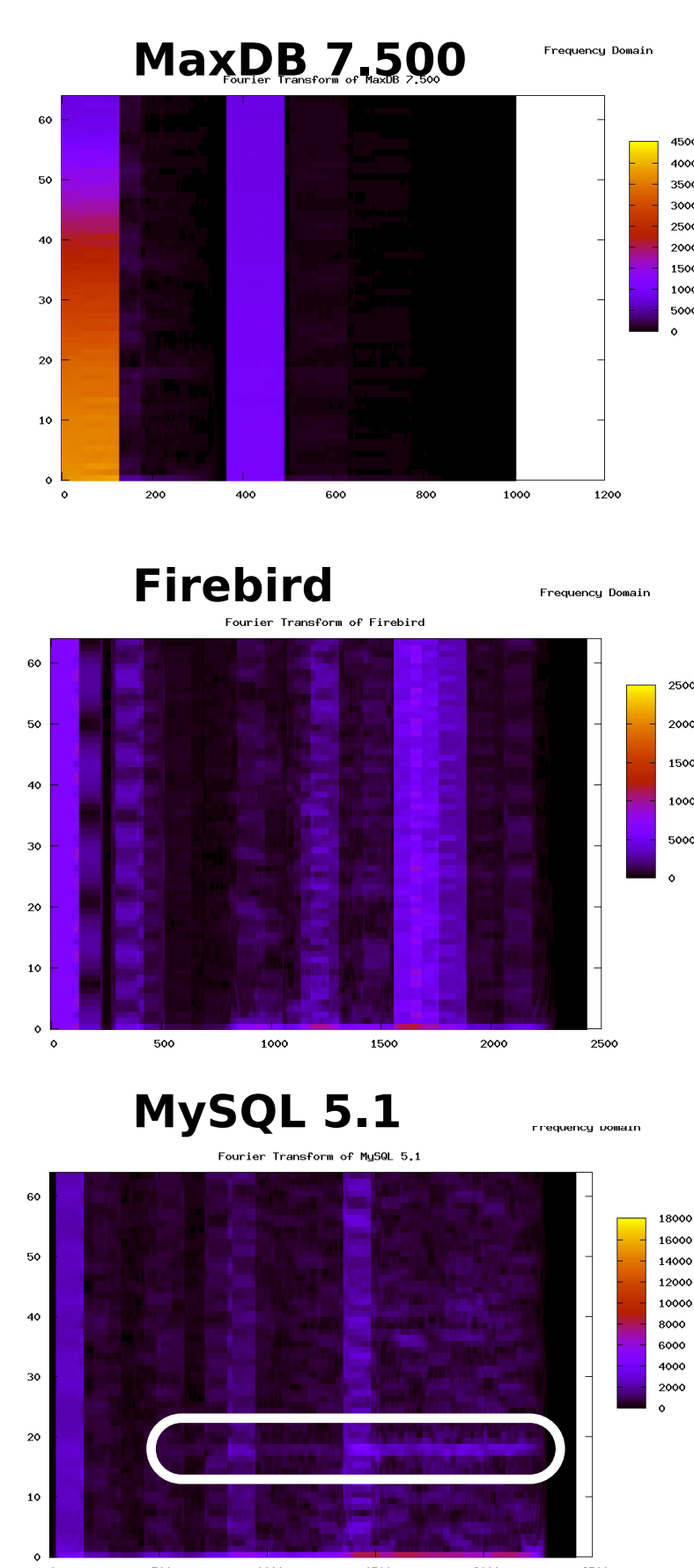
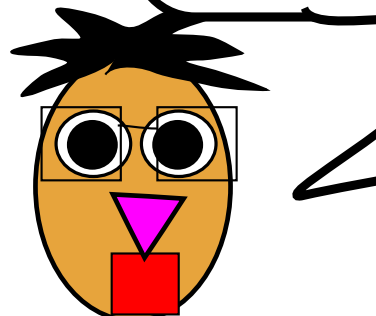
Note the highlighted peaks, these are the dominant frequencies, which you could use for intervals in Time-series analysis.

Spectrograms of software change over time

Spectrograms are Fourier transforms of short periods, shown side by side. The x-axis is time, y-axis frequency and color is magnitude.

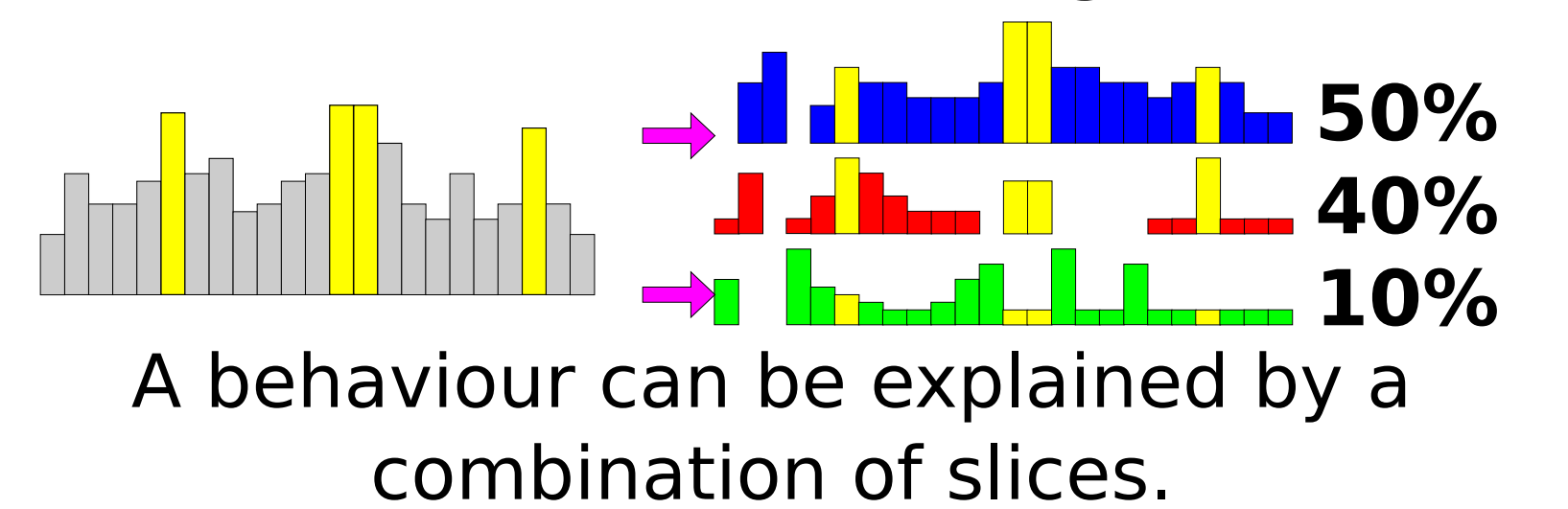


Horizontal smears show recurrent behaviour.

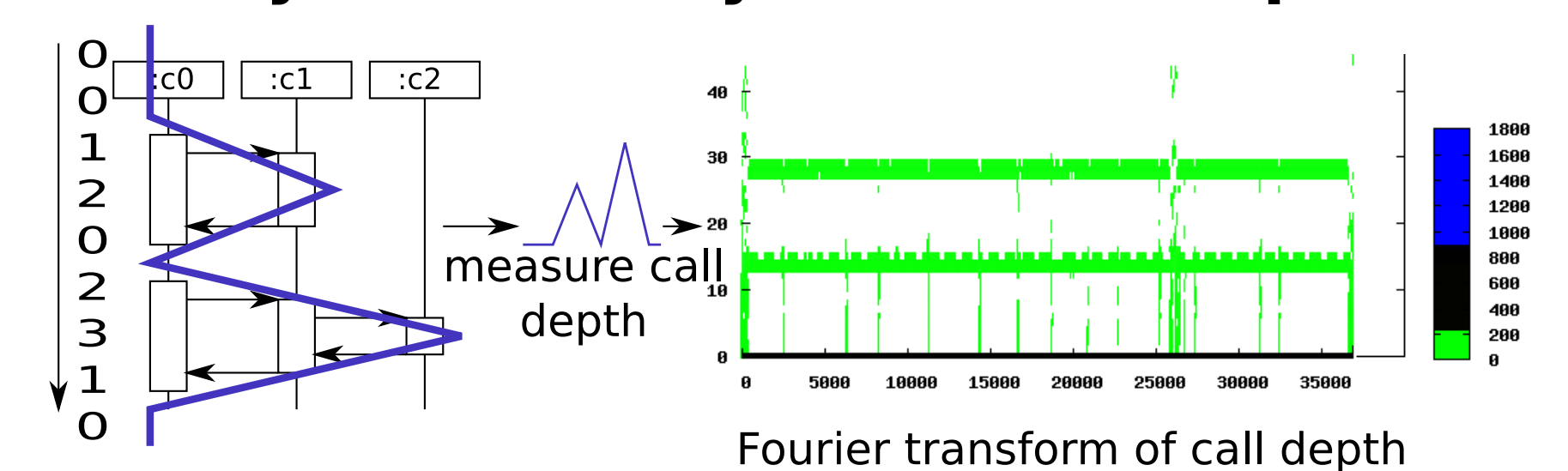


Future Work

Who's responsible for this behaviour? Semantic Slicing



Other uses of the Fourier Transform: Dynamic Analysis via call depth



Partition development time via Self Similarity

