

Composable Statistics

Brian Beckman

May 23, 2019

Contents

1	COMPOSABLE STATISTICS	2
1.1	TODO TANGLE, NOWEB	2
1.2	TODO GRAPHICS	2
1.3	CLOJURE	2
1.4	TODO HOW TO USE THIS DOCUMENT	2
2	INTRODUCTION	2
2.1	TODO: GENERATE NEW RANDOM DATA	3
3	RUNNING COUNT	3
3.1	THREAD-SAFE	3
3.2	AVOIDING <i>REDUCE</i>	4
4	RUNNING MEAN	4
4.1	REMOVING OUTPUT COUPLING	5
4.2	NUMERICAL CHECK	5
5	CORE.ASYNC	6
5.1	SHALLOW TUTORIAL	6
5.2	DEEP TUTORIAL	6
6	ASYNC DATA STREAMS	13
6.1	ASYNC RUNNING MEAN	15
7	RUNNING STDDEV	16
7.1	BRUTE-FORCE (SCALAR VERSION)	16
7.2	DEF Z2S: SMALLER EXAMPLE	17
7.3	REALLY DUMB RECURRENCE	17
7.4	SCHOOL VARIANCE	18
7.5	DEFN MAKE SCHOOL STATS MAPPER	18
7.6	DEFN MAKE RECURRENT STATS MAPPER	19
7.7	DEFN MAKE WELFORD'S STATS MAPPER	19
8	WINDOWED STATISTICS	20
8.1	DEF Z3S: MORE SAMPLE DATA	21
8.2	DEFN MAKE SLIDING STATS MAPPER	21

34	9 KALMAN FILTER	22
35	9.1 BASIC LINEAR ALGEBRA	22
36	9.2 DEFN SYMMETRIC PART	24
37	9.3 DEFN ANTI-SYMMETRIC PART	24
38	9.4 DEFN KALMAN UPDATE: GENERAL EXTENDED KALMAN FILTER	27
39	9.5 DEFN MAKE-KALMAN-MAPPER	29
40	10 VISUALIZATION	30
41	10.1 INCANTER	30
42	10.2 OZ	31
43	11 GAUSSIAN PROCESSES	32
44	11.1 RECURRENT LINEAR REGRESSION	32

45 1 COMPOSABLE STATISTICS

46 1.1 TODO TANGLE, NOWEB

47 This will become a fully literate program via `org-babel tangle` and `noweb`. At present, it is an org file
 48 converted directly from an old iPython (Jupyter) notebook. Jupyter proved not to scale well.

49 1.2 TODO GRAPHICS

50 We still have to work out how to embed graphics in this file.

51 1.3 CLOJURE

52 We prefer Clojure to Python for this exercise due to Clojure's concurrency primitives, especially atoms and
 53 `core.async`. Python is growing and improving rapidly, so we may return to it someday.

54 1.3.1 TODO PROJECT.CLJ

55 At present, the critical file `project.clj` is external to this document. It will be one of the first to tangle.

56 The best sites for learning Clojure by example are clojuredocs.org and 4clojure.org. A recommended
 57 book is Clojure for the Brave and True.

58 1.4 TODO HOW TO USE THIS DOCUMENT

59 Explain how to run Clojure code inside an org-mode buffer, how to tangle and weave, etc.

- 60 1. Install leiningen <https://leiningen.org/> (this is all you need for Clojure)

61 2 INTRODUCTION

62 We want to compute descriptive statistics in constant memory. We want exactly the same code to run over
 63 sequences distributed in space as runs over sequences distributed in time. Sequences distributed in space
 64 are vectors, lists, arrays, lazy or not. Sequences distributed over time are asynchronous streams. Descriptive
 65 statistics range from `count`, `mean`, `max`, `min`, and `variance` to Kalman filters and Gaussian processes. We
 66 decouple computation from data delivery by packaging computation in composable functions.

67 Some sample scalar data:

```
68 (def zs [-0.178654, 0.828305, 0.0592247, -0.0121089, -1.48014,  
69         -0.315044, -0.324796, -0.676357, 0.16301, -0.858164])
```

2.1 TODO: GENERATE NEW RANDOM DATA

3 RUNNING COUNT

The traditional and obvious way with `reduce` and `reductions` (<https://clojuredocs.org/clojure.core/reduce>). *Reduce* takes three arguments: a binary function, an initial value, and a space-sequence of inputs.

```
(reduce
  (fn [count datum] (inc count)) ; binary function
  0                               ; initial value
  zs)                             ; space sequence
```

10

... with all intermediate results:

```
(reductions (fn [c z] (inc c)) 0 zs)

0 1 2 3 4 5 6 7 8 9 10
```

3.1 THREAD-SAFE

Overkill for sequences in space, but safe for multiple threads from asynchronous streams. It also shows (1) *let-over-lambda* (LOL): closing over mutable state variables, and (2) transactional mutation, i.e., *atomic updates*. LOL is semantically equivalent to data encapsulation in OOP, and transactions are easier to verify than is OOP with locks and mutexes.

The following has a defect: we need `initial-count` both to initialize the atom and to initialize the `reduce` call. This defect must be traded off against the generalizable form or *functional type* of the reducible, namely (estimate, measurement) → estimate. We get rid of this defect later.

```
(let [initial-count 0] ; Must use this twice below.
  (reduce
    ; Let-over-lambda (anonymous "object") follows.
    ; "Atom" is a transactional (thread-safe) type in Clojure.
    (let [running-count (atom initial-count)]
      ; That was the "let" of "LOL." Here comes the lambda:
      ; Reducible closure over "running-count."
      (fn [c z] ; Here's the "lambda" of "LOL"
        (swap! running-count inc) ; transactional update
        @running-count))
    initial-count
    zs))
```

10

Showing all intermediate results:

```
(let [initial-count 0]
  (reductions ; <-- this is the only difference to above
    (let [running-count (atom initial-count)]
      (fn [c z]
        (swap! running-count inc)
        @running-count)) ; ~~> new value for c
    initial-count
    zs))
```

0 1 2 3 4 5 6 7 8 9 10

3.2 AVOIDING REDUCE

Reduce only works in space, not in time. Avoiding `reduce` decouples the statistics code (“business logic”) from the space environment (“plumbing”). That space environment delivers data from vectors, lists, etc.). We want to be able to switch out an environment that delivers data from space for an environment that delivers data points z from time.

The following is a thread-safe LOL, without `reduce`. We *map* the LOL over a space-sequence in memory to produce exactly the same result as with `reduce`. The mappable LOL does not need an accumulator argument for `count`.

Below, we map *exactly* the same mappable LOL over asynchronous streams.

A subtle defect: the output is still coupled to the computing environment through `print`. We get rid of that, too, below.

```
(dorun ; <-- Discard 'nil's produced by "print."
  (map
    (let [running-count (atom 0)]
      (fn [z] ; <-- one fewer argument
        (swap! running-count inc)
        (print (str @running-count " "))))
    zs))
```

```
1 2 3 4 5 6 7 8 9 10
```

4 RUNNING MEAN

Consider the following general scheme for recurrence: *a new statistic is an old statistic plus a correction*.

The *correction* is a *gain* times a *residual*. For running mean, the residual is the difference between the new measurement z and the old mean x . The gain is $1/(n+1)$, where n is *count-so-far*. n is a statistic, too, so it is an *old* value, computed and saved before the current observation z arrived.

/The correction therefore depends only on the new input z and on old statistics x and n . The correction does not depend on new statistics/.

Mathematically, write the general recurrence idea without subscripts as

$$x \leftarrow x + K(z - x)$$

or, with Lamport’s notation, wherein new versions of old values get a prime, as an equation

$$x' = x + K(z - x)$$

(z does not have a prime; it is the only exception to the rule that new versions of old quantities have primes).

Contrast the noisy traditional form, which introduces another variable, the index n . This traditional form is objectively more complicated than either of the two above:

$$x_{n+1} = x_n + K(n)(z_{n+1} - x_n)$$

```
(dorun
  (map
    (let [running-stats (atom {:count 0, :mean 0})]
      (fn [z]
        (let [{x :mean, n :count} @running-stats
              n+1 (inc n) ; cool variable name!
              K (/ 1.0 n+1)]
          (swap! running-stats conj
                  [:count n+1
                   [:mean (+ x (* K (- z x)))]))
```

```

157         (println @running-stats)))
158     zs))

159 {:count 1, :mean -0.178654}
160 {:count 2, :mean 0.3248255}
161 {:count 3, :mean 0.2362919}
162 {:count 4, :mean 0.1741917}
163 {:count 5, :mean -0.156674640000000003}
164 {:count 6, :mean -0.18306953333333337}
165 {:count 7, :mean -0.20331617142857145}
166 {:count 8, :mean -0.262446275}
167 {:count 9, :mean -0.21517335555555556}
168 {:count 10, :mean -0.27947242}

```

169 The swap above calls `conj` on the current contents of the atom `running-stats` and on the rest of
 170 the arguments, namely `[:count n+1, :mean ...]`. `conj` is the idiom for “updating” a hashmap, the
 171 hashmap in the atom, the hashmap that starts off as `{:count 0, :mean 0}`.

172 4.1 REMOVING OUTPUT COUPLING

173 Remove `println` from inside the LOL function of `z`. Now the LOL function of `z` is completely decoupled
 174 from its environment. Also, abstract a “factory” method for the LOL, *make-running-stats-mapper*, to clean up
 175 the line that does the printing.

176 4.1.1 MAKE-RUNNING-STATS-MAPPER

```

177 (defn make-running-stats-mapper []
178   (let [running-stats (atom {:count 0 :mean 0 :datum 0})]
179     (fn [z]
180       (let [{x :mean, n :count, _ :datum} @running-stats
181             n+1 (inc n)
182             K (/ 1.0 n+1)]
183         (swap! running-stats conj
184               [:count n+1]
185               [:mean (+ x (* K (- z x)))]
186               [:datum z]))
187         @running-stats)))
188
189 (clojure.pprint/pprint (map (make-running-stats-mapper) zs))

190 ({:count 1, :mean -0.178654, :datum -0.178654}
191  {:count 2, :mean 0.3248255, :datum 0.828305}
192  {:count 3, :mean 0.2362919, :datum 0.0592247}
193  {:count 4, :mean 0.1741917, :datum -0.0121089}
194  {:count 5, :mean -0.156674640000000003, :datum -1.48014}
195  {:count 6, :mean -0.18306953333333337, :datum -0.315044}
196  {:count 7, :mean -0.20331617142857145, :datum -0.324796}
197  {:count 8, :mean -0.262446275, :datum -0.676357}
198  {:count 9, :mean -0.21517335555555556, :datum 0.16301}
199  {:count 10, :mean -0.27947242, :datum -0.858164})

```

200 4.2 NUMERICAL CHECK

201 The last value of the running mean is $-0.279...42$. Check that against an independent calculation.

202 1. DEFN MEAN

```

203     (defn mean [zs] (/ (reduce + zs) (count zs)))
204     (println (mean zs))

205     -0.27947242

```

206 5 CORE.ASYNC

207 For data distributed over time, we'll use Clojure's `core.async`. `Core.async` has some subtleties that we
 208 analyze below.

```

209 (require
210   '[clojure.core.async
211     :refer
212     [sliding-buffer dropping-buffer buffer
213       <!! <!, >!, >!!,
214       go chan onto-chan close!
215       thread alts! alts!! timeout]])

```

216 5.1 SHALLOW TUTORIAL

217 <https://github.com/clojure/core.async/blob/master/examples/walkthrough.clj>

218 5.2 DEEP TUTORIAL

219 The asynchronous, singleton `go` thread is loaded with very lightweight *pseudothreads* (my terminology, not
 220 standard; most things you will read or see about Clojure.async does not carefully distinguish between
 221 threads and pseudothreads, and I think that's not helpful).

222 Pseudothreads are lightweight state machines that pick up where they left off. It is feasible to have
 223 thousands, even millions of them. Pseudothreads don't block, they *park*. *Parking* and *unparking* are very
 224 fast. We can write clean code with pseudothreads because our code looks like it's blocked waiting for input
 225 or blocked waiting for buffer space. Code with blocking I/O is easy to write and to understand. Code in
 226 `go` forms doesn't actually block, just looks like it.

227 Some details are tricky and definitely not easy to divine from the documentation. Hickey's video from
 228 InfoQ 2013 (<https://www.infoq.com/presentations/core-async-clojure>) is more helpful, but
 229 you can only appreciate the fine points after you've stumbled a bit. I stumbled over the fact that buffered
 230 and unbuffered channels have different synchronization semantics. Syntactically, they look the same, but
 231 you cannot, in general, run the same code over an unbuffered channel that works on a buffered channel.
 232 Hickey says this, but doesn't nail it to the mast; doesn't emphasize it with an example, as I do here in this
 233 deep tutorial. He motivates the entire library with the benefits of first-class queues, but fails to emphasize
 234 that, by default, a channel is not a queue but a blocking rendezvous. He does mention it, but one cannot
 235 fully appreciate the ramifications from a passing glance.

236 5.2.1 COMMUNICATING BETWEEN THREADS AND PSEUDOTHEADS

237 Write output to unbuffered channel `c` via `>!` on the asynchronous `go` real-thread and read input from the
 238 same channel `c` via `<!!` on the UI/REPL `println` real-thread. We'll see later that writing via `>!!` to an
 239 unbuffered channel blocks the UI real-thread, so we can't write before reading unbuffered on the UI/REPL
 240 real-thread. However, we can write before reading on a non-blocking pseudothread, and no buffer space is
 241 needed.

```

242 (let [c (chan)]           ;; unbuffered chan
243       (go (>! c 42))       ;; parks if no space in chan
244       (println (<!! c))    ;; blocks UI/REPL until data on c
245       (close! c))         ;; idiom; may be harmless overkill

```

42

In general, single-bang forms work on `go` pseudothreads, and double-bang forms work on real, heavy-weight, Java threads like the UI/REPL thread behind this notebook. In the rest of this notebook, “thread” means “real thread” and we write “pseudothread” explicitly when that’s what we mean.

I don’t address thread leakage carefully in this tutorial, mostly because I don’t yet understand it well. I may overkill by closing channels redundantly.

5.2.2 CHANNEL VODOO FIRST

Writing before reading seems very reasonable, but it does not work on unbuffered channels, as we see below. Before going there, however, let’s understand more corners of the example above.

The `go` form itself returns a channel:

```
(clojure.repl/doc go)

-----
clojure.core.async/go
([& body])
Macro
  Asynchronously executes the body, returning immediately to the
  calling thread. Additionally, any visible calls to <!, >! and alt!/alts!
  channel operations within the body will block (if necessary) by
  ‘parking’ the calling thread rather than tying up an OS thread (or
  the only JS thread when in ClojureScript). Upon completion of the
  operation, the body will be resumed.

  Returns a channel which will receive the result of the body when
  completed
```

I believe “the calling thread” above refers to a pseudothread inside the `go` real-thread, but I am not sure because of the ambiguities in the official documentation between “blocking” and “parking” and between “thread” and “well, we don’t have a name for them, but Brian calls them ‘pseudothreads’.”

Is the channel returned by `go` the same channel as `c`?

```
(let [c (chan)]
  (println {:c-channel c})
  (println {:go-channel (go (>! c 42))})
  (println {:c-coughs-up (<!! c)})
  (println {:close-c (close! c)}))

{:c-channel #object[clojure.core.async.impl.channels.ManyToManyChannel 0x5b941432 clojure.]
 :go-channel #object[clojure.core.async.impl.channels.ManyToManyChannel 0x3c3d4797 clojure.]
 :c-coughs-up 42}
{:close-c nil}
```

No, `c` is a different channel from the one returned by `go`. Consult the documentation for `go` once more:

```
(clojure.repl/doc go)

-----
clojure.core.async/go
([& body])
Macro
  Asynchronously executes the body, returning immediately to the
  calling thread. Additionally, any visible calls to <!, >! and alt!/alts!
  channel operations within the body will block (if necessary) by
```

'parking' the calling thread rather than tying up an OS thread (or the only JS thread when in ClojureScript). Upon completion of the operation, the body will be resumed.

Returns a channel which will receive the result of the body when completed

We should be able to read from the channel returned by `go`; call it `d`:

```
(let [c (chan)
      d (go (>! c 42))] ;; 'let' in Clojure is sequential,
      ;; like 'let*' in Scheme or Common Lisp,
      ;; so 'd' has a value, here.
  (println {:c-coughs-up (<!! c), ;; won't block
            :d-coughs-up (<!! d)}) ;; won't block
  (close! c)
  (close! d))
```

```
{:c-coughs-up 42, :d-coughs-up true}
```

`d's coughing up true` means that the body of the `go`, namely `(>! c 42)` must have returned `true`, because `d` coughs up "the result of the body when completed." Let's see whether our deduction matches documentation for `>!`:

```
(clojure.repl/doc >!)
-----
clojure.core.async/>!
([port val])
  puts a val into port. nil values are not allowed. Must be called
  inside a (go ...) block. Will park if no buffer space is available.
  Returns true unless port is already closed.
```

Sure enough. But something important is true and not obvious from this documentation. Writing to `c` inside the `go` block parks the pseudothread because no buffer space is available: `c` was created with a call to `chan` with no arguments, so no buffer space is allocated. Only when reading from `c` does the pseudothread unpark. How? There is no buffer space. Reading on the UI thread manages to short-circuit any need for a buffer and unpark the pseudothread. Such short-circuiting is called a *rendezvous* in the ancient literature of concurrency. Would the pseudothread unpark if we read inside a `go` block and not on the UI thread?

```
(let [c (chan)
      d (go (>! c 42))
      e (go (<! c))]
  (clojure.pprint/pprint {
    :c-channel c, :d-channel d, :e-channel e,
    :e-coughs-up (<!! e), ;; won't block
    :d-coughs-up (<!! d)}) ;; won't block
  (close! c)
  (close! d)
  (close! e))

{:c-channel
 #object[clojure.core.async.impl.channels.ManyToManyChannel 0x5c01e70 "clojure.core.async.
 :d-channel
 #object[clojure.core.async.impl.channels.ManyToManyChannel 0x749efaf9 "clojure.core.async
 :e-channel
 #object[clojure.core.async.impl.channels.ManyToManyChannel 0x287f4772 "clojure.core.async
 :e-coughs-up 42,
 :d-coughs-up true}
```


Yes, the pseudothread that parked when 42 is put on `c` via `>!` unparks when 42 is taken off via `<!`. Channel `d` represents the parking step and channel `e` represents the unparking step. All three channels are different.

So now we know how to short-circuit or rendezvous unbuffered channels. In fact, the order of reading and writing (taking and putting) does not matter in the nebulous, asynchronous world of pseudothreads. How Einsteinian is that? The following takes (reads) from `c` on `e` before putting (writing) to `c` on `d`. That's the same as above, only in the opposite order.

```
(let [c (chan)
      e (go (<! c))
      d (go (>! c 42))]
  (clojure.pprint/pprint {
    :c-channel c, :d-channel d, :e-channel e,
    :e-coughs-up (<!! e), ;; won't block
    :d-coughs-up (<!! d)})) ;; won't block
(close! c)
(close! d)
(close! e))

{:c-channel
 #object[clojure.core.async.impl.channels.ManyToManyChannel 0x5807392 "clojure.core.async.
 :d-channel
 #object[clojure.core.async.impl.channels.ManyToManyChannel 0x25a60d5b "clojure.core.async
 :e-channel
 #object[clojure.core.async.impl.channels.ManyToManyChannel 0x27805f00 "clojure.core.async
 :e-coughs-up 42,
 :d-coughs-up true}
```

5.2.3 PUTS BEFORE TAKES CONSIDERED RISKY

`>!!`, by default, blocks if called too early on an unbuffered real thread. We saw above that parked pseudothreads don't block: you can read and write to channels in `go` blocks in any order. However, that's not true with threads that actually block. The documentation is obscure, though not incorrect, about this fact.

```
(clojure.repl/doc >!!)

-----
clojure.core.async/>!!
([port val])
  puts a val into port. nil values are not allowed. Will block if no
  buffer space is available. Returns true unless port is already closed.
```

When is “no buffer space available?” It turns out that the default channel constructor makes a channel with no buffer space allocated by default.

```
(clojure.repl/doc chan)

-----
clojure.core.async/chan
([[] [buf-or-n] [buf-or-n xform] [buf-or-n xform ex-handler]])
  Creates a channel with an optional buffer, an optional transducer
  (like (map f), (filter p) etc or a composition thereof), and an
  optional exception-handler. If buf-or-n is a number, will create
  and use a fixed buffer of that size. If a transducer is supplied a
  buffer must be specified. ex-handler must be a fn of one argument -
  if an exception occurs during transformation it will be called with
  the Throwable as an argument, and any non-nil return value will be
  placed in the channel.
```

We can test the blocking-on-unbuffered case as follows. The following code will block at the line (`>!! c 42`), as you'll find if you uncomment the code (remove `#_` at the beginning) and run it. You'll have to interrupt the Kernel using the "Kernel" menu at the top of the notebook, and you might have to restart the Kernel, but you should try it once.

```
#_(let [c (chan)]
      (>!! c 42)
      (println (<!! c))
      (close! c))
```

The following variation works fine because we made "buffer space" before writing to the channel. The only difference to the above is the `1` argument to the call of `chan`.

```
(let [c (chan 1)]
  (>!! c 42)
  (println (<!! c))
  (close! c))
```

42

The difference between the semantics of the prior two examples is not subtle: one hangs the kernel and the other does not. However, the difference in the syntax is subtle and easy to miss.

We can read on the asynchronous `go` pool from the buffered channel `c` because the buffered write (`>!! c`) on the UI thread doesn't block:

```
(let [c (chan 1)]
  (>!! c 42)
  (println {:go-channel-coughs-up (<!! (go (<! c)))}))
  (close! c))

{:go-channel-coughs-up 42}
```

1. ORDER DOESN'T MATTER, SOMETIMES

We can do things backwards, reading before writing, even without a buffer. Read from channel (`<! c`) on the `async go` thread "before" writing to (`>!! c 42`) on the REPL / UI thread. "Before," here, of course, means syntactically or lexically "before," not temporally.

```
(let [c (chan) ;; NO BUFFER!
      d (go (<! c)) ;; park a pseudothread to read c
      e (>!! c 42)] ;; blocking write unparks c's pseudothread
  (println {:c-hangs '(<!! c),
            :d-coughs-up (<!! d),
            :what's-e e})
  (close! c) (close! d))
```

```
{:c-hangs (<!! c), :d-coughs-up 42, :what's-e true}
```

Why did `>!!` produce `true`? Look at docs again:

```
(clojure.repl/doc >!!)
```

```
-----
clojure.core.async/>!!
([port val])
  puts a val into port. nil values are not allowed. Will block if no
  buffer space is available. Returns true unless port is already closed.
```

Ok, now I fault the documentation. `>!!` will block if there is no buffer space available *and* if there is no *rendezvous* available, that is, no pseudothread parked waiting for `<!!`. I have an open question in the Google group for Clojure about this issue with the documentation.

To get the value written in into `c`, we must read `d`. If we tried to read it from `c`, we would block forever because `>!!` blocks when there is no buffer space, and `c` never has buffer space. We get the value out of the `go` nebula by short-circuiting the buffer, by a *rendezvous*, as explained above.

`e`'s being true means that `c` wasn't closed. (`>!! c 42`) should hang.

```
(let [c (chan) ;; NO BUFFER!
      d (go (<! c)) ;; park a pseudothread to read c
      e (>!! c 42) ;; blocking write unparks c's pseudothread
      f '(hangs (>!! c 43))] ;; is 'c' closed?
  (println {:c-coughs-up '(hangs (<!! c)),
            :d-coughs-up (<!! d),
            :what's-e e,
            :what's-f f})
  (close! c) (close! d))
```

```
{:c-coughs-up (hangs (<!! c)), :d-coughs-up 42, :what's-e true, :what's-f (hangs (>!! c 43))}
```

StackOverflow reveals a way to find out whether a channel is closed by peeking under the covers (<https://stackoverflow.com/questions/24912971>):

```
(let [c (chan) ;; NO BUFFER!
      d (go (<! c)) ;; park a pseudothread to read c
      e (>!! c 42) ;; blocking write unparks c's pseudothread
      f (clojure.core.async.impl.protocols/closed? c)]
  (println {:c-coughs-up '(hangs (<!! c)),
            :d-coughs-up (<!! d),
            :c-is-open-at-e? e,
            :c-is-open-at-f? f})
  (close! c) (close! d))
```

```
{:c-coughs-up (hangs (<!! c)), :d-coughs-up 42, :c-is-open-at-e? true, :c-is-open-at-f? false}
```

2. ORDER DOES MATTER, SOMETIMES

Order does matter this time: Writing blocks the UI thread without a buffer and no parked read (*rendezvous*) in the `go` nebula beforehand. I hope you can predict that the following will block even before you run it. To be sure, run it, but you'll have to interrupt the kernel as before.

```
#_(let [c (chan)
        e (>!! c 42) ;; blocks forever
        d (go (<! c))]
  (println {:c-coughs-up '(this will hang (<!! c)),
            :d-coughs-up (<!! d),
            :what's-e e})
  (close! c) (close! d))
```

5.2.4 TIMEOUTS: DON'T BLOCK FOREVER

In all cases, blocking calls like `>!!` to unbuffered channels without timeout must appear *last* on the UI, non-`go`, thread, and then only if there is some parked pseudothread that's waiting to read the channel by short-circuit (*rendezvous*). If we block too early, we won't get to the line that launches the `async go` nebula and parks the short-circuitable pseudothread—parks the *rendezvous*.

The UI thread won't block forever if we add a timeout. `alts!!` is a way to do that. The documentation and examples are difficult, but, loosely quoting (emphasis and edits are mine, major ones in square brackets):

```
(alts!! ports & {:as opts})
```

This destructures all keyword options into `opts`. We don't need `opts` or the `:as` keyword below.

Completes at most one of several channel operations. [/Not for use inside a (go ...) block./] **ports is a vector of channel endpoints**, [A channel endpoint is] either a channel to take from or a vector of [channel-to-put-to val-to-put] pairs, in any combination. Takes will be made as if by `<!!`, and puts will be made as if by `>!!`. If more than one port operation is ready, a non-deterministic choice will be made unless the `:priority` option is true. If no operation is ready and a `:default` value is supplied, [=default-val :default=] will be returned, otherwise `alts!!` will [/block/ xxxpark ?] until the first operation to become ready completes. **Returns [val port] of the completed operation**, where `val` is the value taken for takes, and a boolean (true unless already closed, as per `put`!) for puts. `opts` are passed as `:key val...` Supported options: `:default val` - the value to use if none of the operations are immediately ready `:priority true` - (default nil) when true, the operations will be tried in order. Note: there is no guarantee that the port exprs or val exprs will be used, nor in what order should they be, so they should not be depended upon for side effects.

(alts!! ...) returns a [val port] 2-vector.

(second (alts!! ...)) is a wrapper of channel `c` We can't write to the resulting `timeout` channel because we didn't give it a name.

That's a lot of stuff, but we can divine an idiom: pair a channel `c` that *might* block with a fresh `timeout` channel in an `alts!!`. At most one will complete. If `c` blocks, the `timeout` will cough up. If `c` coughs up before the `timeout` expires, the `timeout` quietly dies (question, is it closed? Will it be left open and leak?)

For a first example, let's make a buffered thread that won't block and pair it with a long timeout. You will see that it's OK to write 43 into this channel (the `[c 43]` term is an implied write; that's clear from the documentation). `c` won't block because it's buffered, it returns immediately, long before the `timeout` could expire.

```
(let [c (chan 1)
      a (alts!! ; outputs a [val port] pair; throw away the val
              ; here are the two channels for `alts!!`
              [[c 43] (timeout 2500)])]
  (clojure.pprint/pprint {:c c, :a a})
  (let [d (go (<! c))]
    (println {:d-returns (<!! d)}))
  (close! c))
```

```
{:c
 #object[clojure.core.async.impl.channels.ManyToManyChannel 0x1f8b3abd "clojure.core.async
 :a
 [true
 #object[clojure.core.async.impl.channels.ManyToManyChannel 0x1f8b3abd "clojure.core.async
 {:d-returns 43}]
```

But, if we take away the buffer, the `timeout` channel wins. The only difference to the above is that instead of creating `c` via `(chan 1)`, that is, with a buffer of length 1, we create it with no buffer (and we quoted out the blocking read of `d` with a tick mark).

```
(let [c (chan)
      a (alts!! ; outputs a [val port] pair; throw away the val
              ; here are the two channels for `alts!!`
              [[c 43] (timeout 2500)])]
```

```

528     (clojure.pprint/pprint {:c c, :a a})
529     (let [d (go (<! c))]
530         (println {:d-is d})
531         '(println {:d-returns (<!! d)})) ;; blocks
532     (close! c))

533 {:c
534  #object[clojure.core.async.impl.channels.ManyToManyChannel 0x1947eeba "clojure.core.async
535  :a
536  [nil
537   #object[clojure.core.async.impl.channels.ManyToManyChannel 0x32a86431 "clojure.core.async
538  {:d-is #object[clojure.core.async.impl.channels.ManyToManyChannel 0x587d8e6d clojure.core.

```

6 ASYNC DATA STREAMS

The following writes at random times (>!) to a parking channel `echo-chan` on an `async go` fast pseudothread. The UI thread block-reads (<!!) some data from `echo-chan`. The UI thread leaves values in the channel and thus leaks the channel according to the documentation for `close!` here <https://clojure.github.io/core.async/api-index.html#C>. To prevent the leak permanently, we close the channel explicitly.

```

545 (def echo-chan (chan))
546
547 (doseq [z zs] (go (Thread/sleep (rand 100)) (>! echo-chan z)))
548 (dotimes [_ 3] (println (<!! echo-chan)))
549
550 (println {:echo-chan-closed?
551          (clojure.core.async.impl.protocols/closed? echo-chan)})
552 (close! echo-chan)
553 (println {:echo-chan-closed?
554          (clojure.core.async.impl.protocols/closed? echo-chan)})

555 0.828305
556 -0.178654
557 -0.324796
558 {:echo-chan-closed? false}
559 {:echo-chan-closed? true}

```

We can chain channels, again with leaks that we explicitly close. Also, we must not >! (send) a nil to `repl-chan`, and <! can produce nil from `echo-chan` after the timeout and we close `echo-chan`.

```

562 (clojure.repl/doc <!)

563 -----
564 clojure.core.async/<!
565 ([port])
566   takes a val from port. Must be called inside a (go ...) block. Will
567   return nil if closed. Will park if nothing is available.

```

Every time you run the block of code below, you will probably get a different result, by design.

```

569 (def echo-chan (chan))
570 (def repl-chan (chan))
571
572 ;; >! chokes on nulls. <! echo-chan can cough up nil if we time out
573 ;; and close the channel. The following line will throw an exception

```

```
574 ;; unless we don't close the channel at the end of this code-block.
575
576 ;; (dotimes [_ 10] (go (>! repl-chan (<! echo-chan))))
577
578 ;; Instead of throwing an exception, just put a random character
579 ;; like \? down the pipe after the echo-chan is closed:
580
581 (dotimes [_ 10] (go (>! repl-chan (or (<! echo-chan) \?))))
582
583 (doseq [z zs] (go (Thread/sleep (rand 100)) (>! echo-chan z)))
584
585 (dotimes [_ 3]
586   (println (<!! (second (alts!! [repl-chan
587                                   (timeout 500)])))))
588
589 ;; Alternatively, we can avoid the exception by NOT closing echo-chan.
590 ;; Not closing echo chan will leak it, and that's a lousy idea.
591
592 (close! echo-chan)
593
594 (close! repl-chan)
595
596 0.16301
597 -0.315044
598 -0.676357
```

598 Reading from echo-chan may hang the UI thread because the UI thread races the internal go thread
599 that reads echo-chan, but the timeout trick works here as above.

```
600 (def echo-chan (chan))
601 (def repl-chan (chan))
602
603 (dotimes [_ 10] (go (>! repl-chan (or (<! echo-chan) \?))))
604 (doseq [z zs] (go (Thread/sleep (rand 100)) (>! echo-chan z)))
605 (dotimes [_ 3]
606   (println (<!! (second (alts!! [echo-chan
607                                   (timeout 500)])))))
608
609 (close! echo-chan)
610 (close! repl-chan)
611
612 nil
613 nil
614 nil
```

614 println on a go pseudoprocess works if we wait long enough. This, of course, is bad practice or “code
615 smell.”

```
616 (def echo-chan (chan))
617
618 (doseq [z zs] (go (Thread/sleep (rand 100)) (>! echo-chan z)))
619 (dotimes [_ 3] (go (println (<! echo-chan))))
620
621 (Thread/sleep 500) ; no visible output if you remove this line.
622 (close! echo-chan)
623
624 -0.324796
625 0.828305
626 -0.676357
```

6.1 ASYNC RUNNING MEAN

6.1.1 DEFN ASYNC-RANDOMIZED-SCAN

We want `running-stats` called at random times and with data in random order. A *transducer*, (`map mapper`), lets us collect items off the buffer. The size of the buffer does not matter, but we must specify it. Notice that the side-effector `effector` is passed in, so `async-randomized-scan` remains decoupled from its environment.

In this style of programming, the asynchronous stream might sometimes be called a *functor*, which is anything that's mappable, anything you can map over.

```
(defn async-randomized-scan [zs mapper effector]
  (let [transducer (map mapper)
        ; give buffer length if there is a transducer
        echo-chan (chan (buffer 1) transducer)]
    (doseq [z zs]
      (go (Thread/sleep (rand 100)) (>! echo-chan z)))
    (dotimes [_ (count zs)] (effector (<!! echo-chan)))
    (close! echo-chan)))

(async-randomized-scan zs (make-running-stats-mapper) println)

{:count 1, :mean 0.0592247, :datum 0.0592247}
{:count 2, :mean -0.05971465, :datum -0.178654}
{:count 3, :mean 0.014526899999999998, :datum 0.16301}
{:count 4, :mean -0.35913982499999997, :datum -1.48014}
{:count 5, :mean -0.35032065999999995, :datum -0.315044}
{:count 6, :mean -0.40466004999999994, :datum -0.676357}
{:count 7, :mean -0.39325089999999996, :datum -0.324796}
{:count 8, :mean -0.34560814999999995, :datum -0.0121089}
{:count 9, :mean -0.21517335555555555, :datum 0.828305}
{:count 10, :mean -0.27947241999999994, :datum -0.858164}
```

We don't need to explicitly say `buffer`, but I prefer to do.

6.1.2 DEFN MAKE SOW REAP

The `effector` above just prints to the console. Suppose we want to save the data?

The following is a version of Wolfram's `Sow` and `Reap` that does not include tags. It uses `atom` for an effectful store because a `let` variable like `result` is not a `var` and `alter-var-root` won't work on (`let [result []] ...`). An `atom` might be overkill.

`make-sow-reap` returns a message dispatcher in the style of *The Little Schemer*. It responds to namespaced keywords `::sow` and `::reap`. In the case of `::sow`, it returns an `effector` function that `conj`'s its input to the internal result atomically. In the case of `::reap`, it returns the value of the result accumulated so-far.

```
(do (defn make-sow-reap []
      (let [result (atom [])]
        (fn [msg]
          (cond
            (identical? msg ::sow)
            (fn [x] (swap! result #(conj % x)))
            (identical? msg ::reap)
            @result))))))

(let [accumulator (make-sow-reap)]
  (async-randomized-scan zs
```

```

675                                     (make-running-stats-mapper)
676                                     (accumulator ::sow))
677      (last (accumulator ::reap)))

678      :count 10 :mean -0.27947242 :datum -1.48014

```

Occasionally, there is some floating-point noise in the very low digits of the mean because `async-randomized-scan` scrambles the order of the inputs. The mean should always be almost equal to -0.27947242 .

6.1.3 DEFN ASYNC NON RANDOM SCAN

Of course, the mean of any permutation of the data `zs` is the same, so the order in which data arrive does not change the final result, except for some occasional floating-point noise as mentioned above.

```

684 (do (defn async-non-random-scan [zs mapper effector]
685     (let [transducer (map mapper)
686           echo-chan (chan (buffer 1) transducer)]
687       (go (doseq [z zs] (>! echo-chan z)))
688       (dotimes [_ (count zs)] (effector (<!! echo-chan)))
689       (close! echo-chan)))
690
691     (let [accumulator (make-sow-reap)]
692       (async-non-random-scan zs (make-running-stats-mapper)
693                               (accumulator ::sow))
694       (last (accumulator ::reap)))

695     :count 10 :mean -0.27947242 :datum -0.858164

```

6.1.4 DEFN SYNC SCAN: WITH TRANSDUCER

Here is the modern way, with `transduce`, to reduce over a sequence of data, in order. It's equivalent to the non-random `async` version above. The documentation for `transduce` writes its parameters as `xform f coll`, and then says

reduce with a transformation of `f (xf)`. If `init` is not supplied, `(f)` will be called to produce it.

Our `xform` is `transducer`, or `(map mapper)`, and our `f` is `conj`, so this is an idiom for mapping because `(conj)`, with no arguments, returns `[]`, an appropriate `init`.

```

704 (do (defn sync-scan [zs mapper]
705     (let [transducer (map mapper)]
706       (transduce transducer conj zs)))
707
708     (last (sync-scan zs (make-running-stats-mapper)))

709     :count 10 :mean -0.27947242 :datum -0.858164

```

We now have complete symmetry between space and time, space represented by the vector `zs` and time represented by values on `echo-chan` in random and in non-random order.

7 RUNNING STDDEV

7.1 BRUTE-FORCE (SCALAR VERSION)

The definition of variance is the following, for $N > 1$:

$$\frac{1}{N-1} \sum_{i=1}^N (z_i - \bar{z}_N)^2$$

715 The sum is the *sum of squared residuals*. Each residual is the difference between the i -th datum z_i and
 716 the mean \bar{z}_N of all N data in the sample. The outer constant, $1/(N-1)$ is Bessel's correction.

717 7.1.1 DEFN SSR: SUM OF SQUARED RESIDUALS

718 The following is *brute-force* in the sense that it requires all data up-front so that it can calculate the mean.

```
719 (do (defn ssr [sequ]
720     (let [m (mean sequ)]
721         (reduce #(+ %1 (* (- %2 m) (- %2 m)))
722                 0 sequ)))
723     (ssr zs) )
724 3.5566483654807355
```

725 7.1.2 DEFN VARIANCE

726 Call `ssr` to compute variance:

```
727 (do
728 (defn variance [sequ]
729     (let [n (count sequ)]
730         (case n
731             0 0
732             1 (first sequ)
733             #_default (/ (ssr sequ) (- n 1.0)))))
734 (variance zs) )
735 0.3951831517200817
```

736 7.2 DEF Z2S: SMALLER EXAMPLE

737 Let's do a smaller example:

```
738 (do (def z2s [55. 89. 144.])
739     (variance z2s) )
740 2017.0
```

741 7.3 REALLY DUMB RECURRENCE

742 Remember our general form for recurrences, $x \leftarrow x + K \times (z - x)$?

743 We can squeeze running variance into this form in a really dumb way. The following is really dumb
 744 because:

- 745 1. it requires the whole sequence up front, so it doesn't run in constant memory
- 746 2. the intermediate values are meaningless because they refer to the final mean and count, not to the
 747 intermediate ones

748 But, the final value is correct.

```

749 (do (reductions
750     (let [m (mean z2s) ; uh-oh, we refer to _all_ the data ??
751           c (count z2s)]
752       (fn [var z] (+ var (let [r (- z m)] ; residual
753                             (/ (* r r) (- c 1.0))))))
754     0 z2s) )

```

755 That was so dumb that we won't bother with a thread-safe, stateful, or asynchronous form.

756 7.4 SCHOOL VARIANCE

757 For an easy, school-level exercise, prove the following equation:

$$\frac{1}{N-1} \sum_{i=1}^N (z_i - \bar{z}_N)^2 = \frac{1}{N-1} \left(\sum_{i=1}^N (z_i^2) - N \bar{z}_N^2 \right)$$

758 Instead of the sum of squared residuals, *ssr*, accumulate the sum of squares, *ssq*.

759 *School variance* is exposed to *catastrophic cancellation* because *ssq* grows quickly. We fix that defect below.

760 We see that something is not best with this form because we don't use the old variance to compute the
761 new variance. We do better below.

762 Of course, the same mapper works synchronously and asynchronously.

763 7.5 DEFN MAKE SCHOOL STATS MAPPER

764 and test it both synchronously and asynchronously, randomized and not:

```

765 (defn make-school-stats-mapper []
766   (let [running-stats (atom {:count 0, :mean 0,
767                             :variance 0, :ssq 0})]
768     (fn [z]
769       (let [{x :mean, n :count, s :ssq} @running-stats
770             n+1 (inc n)
771             K    (/ 1.0 n+1)
772             r    (- z x)
773             x'   (+ x (* K r)) ;; Isn't prime notation nice?
774             s'   (+ s (* z z))]
775         (swap! running-stats conj
776               [:count      n+1]
777               [:mean       x']
778               [:ssq        s']
779               [:variance   (/ (- s' (* n+1 x' x')) (max 1 n))]))
780       @running-stats)))
781
782 (clojure.pprint/pprint (sync-scan z2s (make-school-stats-mapper)))
783
784 (async-randomized-scan z2s (make-school-stats-mapper) println)
785
786 (async-non-random-scan z2s (make-school-stats-mapper) println)
787
788 [{:count 1, :mean 55.0, :variance 0.0, :ssq 3025.0}
789  {:count 2, :mean 72.0, :variance 578.0, :ssq 10946.0}
790  {:count 3, :mean 96.0, :variance 2017.0, :ssq 31682.0}]
791
792 [{:count 1, :mean 55.0, :variance 0.0, :ssq 3025.0}
793  {:count 2, :mean 99.5, :variance 3960.5, :ssq 23761.0}
794  {:count 3, :mean 96.0, :variance 2017.0, :ssq 31682.0}
795  {:count 1, :mean 55.0, :variance 0.0, :ssq 3025.0}]

```

```

794 {:count 2, :mean 72.0, :variance 578.0, :ssq 10946.0}
795 {:count 3, :mean 96.0, :variance 2017.0, :ssq 31682.0}

```

7.6 DEFN MAKE RECURRENT STATS MAPPER

We already know the recurrence for the mean:

$$x \leftarrow x + K \cdot (z - x) = x + \frac{1}{n+1}(z - x)$$

We want a recurrence with a similar form for the variance. It takes a little work to prove, but it's still a school-level exercise. K remains $1/(n+1)$, the value needed for the new mean. We could define a pair of gains, one for the mean and one for the variance, but it would be less pretty.

$$v \leftarrow \frac{(n-1)v + K n (z-x)^2}{\max(1, n)}$$

```

801 (defn make-recurrent-stats-mapper []
802   (let [running-stats (atom {:count 0, :mean 0,
803                               :variance 0})]
804     (fn [z]
805       (let [{x :mean, n :count, v :variance} @running-stats
806             n+1 (inc n)
807             K (/ 1.0 (inc n))
808             r (- z x)
809             x' (+ x (* K r))
810             ssr (+ (* (- n 1) v) ; old ssr is (* (- n 1) v)
811                   (* K n r r))]
812         (swap! running-stats conj
813               [:count n+1]
814               [:mean x']
815               [:variance (/ ssr (max 1 n))]))
816       @running-stats)))
817
818 (async-non-random-scan z2s (make-recurrent-stats-mapper) println)
819
819 {:count 1, :mean 55.0, :variance 0.0}
820 {:count 2, :mean 72.0, :variance 578.0}
821 {:count 3, :mean 96.0, :variance 2017.0}

```

7.7 DEFN MAKE WELFORD'S STATS MAPPER

The above is equivalent, algebraically and numerically, to Welford's famous recurrence for the sum of squared residuals S . In recurrences, we want everything on the right-hand sides of equations or left arrows to be old, *prior* statistics, except for the new observation / measurement / input z . Welford's requires the new, *posterior* mean on the right-hand side, so it's not as elegant as our recurrence above. However, it is easier to remember!

$$S \leftarrow S + (z - x_N)(z - x_{N+1}) = S + (z - x)(z - (x + K(z - x)))$$

```

828 (do (defn make-welfords-stats-mapper []
829     (let [running-stats (atom {:count 0, :mean 0, :variance 0})]
830       (fn [z]
831         (let [{x :mean, n :count, v :variance} @running-stats
832               n+1 (inc n)
833               K (/ 1.0 n+1)
834               r (- z x)

```

```

835         x' (+ x (* K r))
836         ssr (+ (* (- n 1) v)
837               ;; only difference to recurrent variance:
838               (* (- z x) (- z x')))]
839         (swap! running-stats conj
840              [:count      n+1]
841              [:mean       x' ]
842              [:variance (/ ssr (max 1 n))]))
843         @running-stats)))
844
845     (async-non-random-scan
846       z2s (make-welfords-stats-mapper) println) )
847
848     {:count 1, :mean 55.0, :variance 0.0}
849     {:count 2, :mean 72.0, :variance 578.0}
850     {:count 3, :mean 96.0, :variance 2017.0}

```

8 WINDOWED STATISTICS

Suppose we want running statistics over a history of fixed, finite length. For example, suppose we have $N = 10$ data and we want the statistics in a window of length $w = 3$ behind the current value, inclusively. When the first datum arrives, the window and the total include one datum. The window overhangs the left until the third datum. When the fourth datum arrives, the window contains three data and the total contains four data. After the tenth datum, we may consider three more steps marching the window “off the cliff” to the right. The following figure illustrates (the first row corresponds to $n = 0$, not to $n = 1$):

We won’t derive the following formulas, but rather say that they have been vetted at least twice independently (in a C program and in a Mathematica program). The following table shows a unit test that we reproduce. The notation is explained after the table.

Denote prior statistics by plain variables like m and corresponding posteriors by the same variables with primes like m' . The posteriors j and u do not have a prime.

variable	description
n	prior count of data points; equals 0 when considering the first point
z	current data point
w	fixed, constant, maximum width of window; $w \geq 1$
j	posterior number of points left of the window; $j \geq 0$
u	posterior number of points including z in the running window; $1 \leq u \leq w$
m	prior mean of all points, not including z
m'	posterior mean of all points including z
m_j	prior mean of points left of the window, lagging w behind m
m'_j	posterior mean of points left of the window
m'_w	posterior mean of points in the window, including the current point z
v	prior variance, not including z
v'	posterior variance of all points including z
v_j	prior variance of points left of the window, lagging w behind u_n
v'_j	posterior variance of points left of the window
v'_w	posterior variance of points within the window

The recurrences for m , v , m_j , and v_j have only priors (no primes) on their right-hand sides. The values of m_w and v_w are not recurrences because the non-primed versions do not appear on the right-hand sides of equations 10 and 13. Those equations are simply transformations of the posteriors (values with primes) m' , m'_j , v' , and v'_j .

$$\begin{aligned}
j &= \max(0, n + 1 - w) \\
u &= n - j + 1 \\
m' &= m + \frac{z - m}{n + 1} \\
m'_j &= \begin{cases} m_j + \frac{z_j - m_j}{j} & j > 0 \\ 0 & \text{otherwise} \end{cases} \\
m'_w &= \frac{(n + 1) m' - j m'_j}{u} \\
v' &= \frac{(n - 1) v + \frac{n}{n + 1} (z - m)^2}{\max(1, n)} \\
v'_j &= \begin{cases} \frac{j - 2}{j - 1} v_j + \frac{1}{j} (z_j - m_j)^2 & j > 1 \\ 0 & \text{otherwise} \end{cases} \\
v'_w &= \frac{n v' + (n - w) v'_j + (n + 1) m'^2 - j m_j'^2 - u m_w'^2}{\max(1, u - 1)}
\end{aligned}$$

867 Here is sample data we can compare with the unit test above.

8.1 DEF Z3S: MORE SAMPLE DATA

```
869 (def z3s [0.857454, 0.312454, 0.705325, 0.8393630, 1.637810,
870          0.699257, -0.340016, -0.213596, -0.0418609, 0.054705])
```

871 The best algorithm we have found for tracking historical data is to keep a FIFO queue in a Clojure *vector*
872 of length *w*. This is still constant memory because it depends only on the length *w* of the window, not on
873 the length of the data stream.

8.1.1 DEFN PUSH TO BACK

```
875 (defn push-to-back [item vek]
876   (conj (vec (drop 1 vek)) item))
```

8.2 DEFN MAKE SLIDING STATS MAPPER

```
878 (defn make-sliding-stats-mapper [w]
879   (let [running-stats (atom {:n 0, :m 0, :v 0,
880                               :win (vec (repeat w 0)),
881                               :mw 0, :vw 0,
882                               :mj 0, :vj 0})]
883     (fn [z]
884       (let [{:keys [m n v win mj vj]} @running-stats
885             zj (first win)
886             win' (push-to-back z win)
887             n+1 (double (inc n))
888             n-1 (double (dec n))
889             K (/ 1.0 n+1)
890             Kv (* n K)
891             r (- z m)
892             j (max 0, (- n+1 w))
893             u (- n+1 j)
894             m' (+ m (* K r))
```

```

895         rj    (- zj mj)
896         mj'   (if (> j 0), (+ mj (/ rj j)), 0)
897         mw'   (/ (- (* n+1 m') (* j mj')) u)
898         v'    (/ (+ (* n-1 v) (* Kv r r))
899                (max 1 n))
900         vj'   (if (> j 1)
901                (let [j21 (/ (- j 2.0)
902                             (- j 1.0))]
903                  (+ (* j21 vj)
904                    (/ (* rj rj) j)))
905                0)
906         vw'   (let [t1 (- (* n v')
907                          (* (- n w) vj'))
908                  t2 (- (* n+1 m' m')
909                          (* j mj' mj'))
910                  t3 (- (* u mw' mw'))]
911                (/ (+ t1 t2 t3)
912                   (max 1 (- u 1))))
913     ]
914     (swap! running-stats conj
915          [:n      n+1 ]
916          [:m      m'   ]
917          [:v      v'   ]
918          [:mj     mj'  ]
919          [:vj     vj'  ]
920          [:mw     mw'  ]
921          [:vw     vw'  ]
922          [:win    win' ]))
923     @running-stats)))
924
925 (clojure.pprint/print-table
926   [:n :mw :vw]
927   (sync-scan z3s (make-sliding-stats-mapper 3)))

```

```

928
929 |   :n |                               :mw |                               :vw |
930 |-----+-----+-----+-----|
931 |  1.0 |                0.857454 |                0.0 |
932 |  2.0 |                0.584954 |  0.14851250000000005 |
933 |  3.0 |  0.6250776666666666 |  0.07908597588033339 |
934 |  4.0 |  0.6190473333333332 |  0.07499115039433346 |
935 |  5.0 |  1.0608326666666668 |  0.2541686787463333 |
936 |  6.0 |                1.05881 |  0.25633817280899995 |
937 |  7.0 |  0.6656836666666668 |  0.9787942981023336 |
938 |  8.0 |  0.04854833333333334 |  0.3215618307563336 |
939 |  9.0 | -0.19849096666666663 |  0.022395237438003604 |
940 | 10.0 | -0.06691730000000007 |  0.01846722403596973 |

```

941 ... passing the unit test.

942 9 KALMAN FILTER

943 9.1 BASIC LINEAR ALGEBRA

944 Go for high performance with CUDA or Intel KML later.

945 Add the following lines to `project.clj` in the directory that contains this org file:

946 9.1.1 TODO: FULLY LITERATE: TANGLE PROJECT.CLJ

```
947 [net.mikera/core.matrix "0.62.0"]
948 [net.mikera/vectorz-clj "0.48.0"]
949 [org.clojure/algolib "0.1.2"]
```

950 Smoke test:

```
951 (require '[clojure.core.matrix :as ccm])
952 (ccm/set-current-implementation :vectorz)
```

```
953 (ccm/shape
954   (ccm/array [[1 2 3]
955               [1 3 8]
956               [2 7 4]]))
```

957 3 3

958 Bits and pieces we will need:

```
959 (ccm/transpose
960   (ccm/array [[1 2 3]
961               [1 3 8]
962               [2 7 4]]))
```

```
963 #vectorz/matrix [[1.0,1.0,2.0],
964                  [2.0,3.0,7.0],
965                  [3.0,8.0,4.0]]
```

966 `mmul` is multiadic (takes more than two arguments). This is possible because matrix multiplication is associative.

```
968 (let [A (ccm/array [[1 2 3]
969                    [1 3 8]
970                    [2 7 4]])]
971   (ccm/mmul (ccm/transpose A) A (ccm/inverse A)))
```

```
972 #vectorz/matrix [[1.0000000000000003,1.0,2.0000000000000004],
973                  [2.00000000000000093,3.0000000000000001,6.999999999999998],
974                  [3.0000000000000006,8.0,3.9999999999999999]]
```

975 9.1.2 DEFN Linspace

```
976 (defn linspace
977   "A sequence of $n$ equally spaced points in the doubly closed
978   interval $[a,b]$, that is, inclusive of both ends."
979   [a b n]
980   (let [d (/ (- b a) (dec n))]
981     (map (fn [x] (+ a (* x d))) (range n))))
982 (clojure.pprint/pprint (linspace 2 3. 3))
983 (2.0 2.5 3.0)
```

9.2 DEFN SYMMETRIC PART

```
(do (defn symmetric-part [M]
      (ccm/div (ccm/add M (ccm/transpose M)) 2.0))
    (symmetric-part [[1 2 3]
                     [1 3 8]
                     [2 7 4]]))
```

1.0	1.5	2.5
1.5	3.0	7.5
2.5	7.5	4.0

9.3 DEFN ANTI-SYMMETRIC PART

```
(do (defn anti-symmetric-part [M]
      (ccm/div (ccm/sub M (ccm/transpose M)) 2.0))
    (anti-symmetric-part [[1 2 3]
                           [1 3 8]
                           [2 7 4]]))
```

0.0	0.5	0.5
-0.5	0.0	0.5
-0.5	-0.5	0.0

```
(let [M [[1 2 3]
         [1 3 8]
         [2 7 4]]]
      (ccm/sub (ccm/add (symmetric-part M)
                        (anti-symmetric-part M))
                M))
```

0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0

9.3.1 DEFN MATRIX ALMOST =

```
(require '[clojure.algo.generic.math-functions :as gmf])
```

The following isn't the best solution: neither relative nor absolute differences are robust. Units in Last Place (ULP) are a better criterion, however, this will unblock us for now.

```
(do (defn matrix-almost=
      ([m1 m2 eps]
       "Checks for near equality against a given absolute difference."
       (mapv (fn [row1 row2]
                (mapv (fn [e1 e2] (gmf/approx= e1 e2 eps))
                      row1 row2))
              m1 m2))
      ([m1 m2]
       "Checks for near equality against a default absolute difference of 1.0e-9"
       (matrix-almost= m1 m2 1.0e-9)))

    (let [M [[1 2 3]
              [1 3 8]
              [2 7 4]]]
      (matrix-almost= (ccm/add (symmetric-part M)
                                (anti-symmetric-part M))
                      M)))
```



```

true true true
true true true
true true true

```

9.3.2 DEFN SIMILARITY TRANSFORM

```

(defn similarity-transform [A M]
  (ccm/mmul A M (ccm/transpose A)))

```

9.3.3 VECTORS, ROW VECTORS, COLUMN VECTORS

The library (like many others) is loose about matrices times vectors.

```

(ccm/mmul
  (ccm/matrix [[1 2 3]
               [1 3 8]
               [2 7 4]]))
(ccm/array [22 23 42]))

#vectorz/vector [194.0, 427.0, 373.0]

```

Pedantically, a matrix should only be allowed to left-multiply a column vector, i.e., a 1×3 matrix. The Clojure library handles this case.

```

(ccm/mmul
  (ccm/matrix [[1 2 3]
               [1 3 8]
               [2 7 4]]))
(ccm/array [[22] [23] [42]]))

#vectorz/matrix [[194.0],
                 [427.0],
                 [373.0]]

```

Non-pedantic multiplication of a vector on the right by a matrix:

```

(ccm/mmul
  (ccm/array [22 23 42])
  (ccm/matrix [[1 2 3]
               [1 3 8]
               [2 7 4]]))

#vectorz/vector [129.0, 407.0, 418.0]

```

Pedantic multiplication of a row vector on the right by a matrix:

```

(ccm/mmul
  (ccm/array [[22 23 42]])
  (ccm/matrix [[1 2 3]
               [1 3 8]
               [2 7 4]]))

#vectorz/matrix [[129.0, 407.0, 418.0]]

```

9.3.4 SOLVING INSTEAD OF INVERTING

Textbooks will tell you that, if you have $Ax = b$ and you want x , you should compute $A^{-1}b$. Don't do this; the inverse is numerically risky and almost never needed:

```
(ccm/mmul
  (ccm/inverse
    (ccm/array [[1 2 3]
                 [1 3 8]
                 [2 7 4]]))
  (ccm/array [22 23 42]))

#vectorz/vector [22.05882352941177,-0.4705882352941142,0.2941176470588234]
```

Instead, use a linear solver. Almost everywhere that you see $A^{-1}b$, visualize `solve(A, b)`. You will get a more stable answer. Notice the difference in the low-significance digits below. The following is a more reliable answer:

```
(require '[clojure.core.matrix.linear :as ccml])

(ccml/solve
  (ccm/array [[1 2 3]
               [1 3 8]
               [2 7 4]]))
  (ccm/array [22 23 42]))

(ccml/solve
  (ccm/matrix [[1 2 3]
                [1 3 8]
                [2 7 4]]))
  (ccm/matrix [22 23 42]))

#vectorz/vector [22.058823529411764,-0.4705882352941176,0.2941176470588236]

(ccm/shape (ccm/matrix [[22] [23] [42]]))

3 1
```

9.3.5 DEFN SOLVE MATRIX

We need `solve` to work on matrices:

```
(defn solve-matrix
  "The 'solve' routine in clojure.core.matrix only works on Matrix times Vector.
  We need it to work on Matrix times Matrix. The equation to solve is
  Ann * Xnm = Bnm
  Think of the right-hand side matrix Bnm as a sequence of columns. Iterate over
  its transpose, treating each column as a row, then converting that row to a
  vector, to get the transpose of the solution X."
  [Ann Bnm]
  (ccm/transpose (mapv (partial ccml/solve Ann) (ccm/transpose Bnm))))

(solve-matrix
  (ccm/matrix [[1 2 3]
                [1 3 8]
                [2 7 4]]))
  (ccm/matrix [[22] [23] [42]]))
```

22.058823529411764
 -0.4705882352941176
 0.2941176470588236

```
1108 (solve-matrix
1109   (ccm/matrix [[1 2 3]
1110               [1 3 8]
1111               [2 7 4]]))
1112   (ccm/matrix [[22 44]
1113               [23 46]
1114               [42 84]]))
```

```
1115 22.058823529411764 44.11764705882353
-0.4705882352941176 -0.9411764705882352
0.2941176470588236 0.5882352941176472
```

1116 9.4 DEFN KALMAN UPDATE: GENERAL EXTENDED KALMAN FILTER

1117 Use Clojure's destructuring to write the Kalman filter as a binary function. See <http://vixra.org/abs/1606.0348>

1119 $\mathbf{x}_{n,1}$ denotes a vector \mathbf{x} with dimension $n \times 1$, that is, a column vector of height n . $\mathbf{P}_{n,n}$ denotes a covariance matrix of dimension $n \times n$, and so on.

1121 The math is as follows (notice step 6 has the same form as all earlier statistics calculations in this document):

1123 Letting inputs:

- 1124 • $\mathbf{x}_{n,1}$ be the current, best estimate of the n -dimensional state of a system
- 1125 • $\mathbf{P}_{n,n}$ be the current, best estimate of the $n \times n$ covariance of state $\mathbf{x}_{n,1}$
- 1126 • $\mathbf{z}_{m,1}$ be the current, m -dimensional observation
- 1127 • $\mathbf{H}_{m,n}$ be linearized observation model to be inverted: $\mathbf{z}_{m,1} = \mathbf{H}_{m,n} \cdot \mathbf{x}_{n,1}$
- 1128 • $\mathbf{A}_{n,n}$ be linearized dynamics
- 1129 • $\mathbf{Q}_{n,n}$ be process noise (covariance) accounting for uncertainty in $\mathbf{A}_{n,n}$
- 1130 • $\mathbf{R}_{m,m}$ be observation noise (covariance) accounting for uncertainty in $\mathbf{z}_{m,1}$

1131 and intermediates and outputs:

- 1132 • $\mathbf{x}'_{n,1}$ (intermediate; *update*) be the estimate of the state after enduring one time step of linearized dynamics
- 1134 • $\mathbf{x}''_{n,1}$ (output; *prediction*) be the estimate of the state after dynamics and after information from the observation $\mathbf{z}_{m,1}$
- 1136 • $\mathbf{P}'_{n,n}$ (intermediate; *update*) be the current, best estimate of the $n \times n$ covariance of state $\mathbf{x}_{n,1}$ after dynamics
- 1138 • $\mathbf{P}''_{n,n}$ (output; *prediction*) be the current, best estimate of the $n \times n$ covariance of state $\mathbf{x}_{n,1}$ after dynamics and observation $\mathbf{z}_{m,1}$

1140 The steps are:

- 1141 1. *Update state estimate*: $\mathbf{x}'_{n,1} = \mathbf{A}_{n,n} \mathbf{x}_{n,1}$
- 1142 2. *Update state covariance*: $\mathbf{P}'_{n,n} = \mathbf{Q}_{n,n} + (\mathbf{A}_{n,n} \mathbf{P}_{n,n} \mathbf{A}_{n,n}^T)$
- 1143 3. *Covariance-update scaling matrix*: $\mathbf{D}_{m,m} = \mathbf{R}_{m,m} + (\mathbf{H}_{m,n} \mathbf{P}'_{n,n} \mathbf{H}_{m,n}^T)$

4. *Kalman gain*: $\mathbf{K}_{n,m} = \mathbf{P}_{n,n} \mathbf{H}_{m,n}^T \mathbf{D}_{m,m}^{-1}$

(a) written as $\mathbf{K}_{n,m}^T = \text{solve}(\mathbf{D}_{m,m}^T, \mathbf{H}_{m,n} \mathbf{P}_{n,n}^T)$

5. *Innovation: predicted observation residual*: $\mathbf{r}_{m,1} = \mathbf{z}_{m,1} - \mathbf{H}_{m,n} \mathbf{x}'_{n,1}$

6. *State prediction*: $\mathbf{x}''_{n,1} = \mathbf{x}'_{n,1} + \mathbf{K}_{n,m} \mathbf{r}_{m,1}$

7. *Covariance reduction matrix*: $\mathbf{L}_{n,n} = \mathbf{I}_{n,n} - \mathbf{K}_{n,m} \mathbf{H}_{m,n}$

8. *Covariance prediction*: $\mathbf{P}'_{n,n} = \mathbf{L}_{n,n} \mathbf{P}'_{n,n}$

```
(defn kalman-update [{:keys [xn1 Pnn]} {:keys [zm1 Hmn Ann Qnn Rmm]})
  (let [x'nl (ccm/mmul Ann xn1) ; Predict state
        P'nn (ccm/add
              Qnn (similarity-transform Ann Pnn)) ; Predict covariance
        Dmm (ccm/add
              Rmm (similarity-transform Hmn P'nn)) ; Gain precursor
        DTmm (ccm/transpose Dmm) ; Support for "solve"
        HP'Tmn (ccm/mmul Hmn (ccm/transpose P'nn)) ; Support for "solve"
        ; Eqn 3 of http://vixra.org/abs/1606.0328:
        KTmn (solve-matrix DTmm HP'Tmn)
        Knm (ccm/transpose KTmn) ; Kalman gain
        ; innovation = predicted obn residual
        rml (ccm/sub zm1 (ccm/mmul Hmn x'nl))
        x''nl (ccm/add x'nl (ccm/mmul Knm rml)) ; final corrected estimate
        n (ccm/dimension-count xn1 0)
        ; new covariance ? catastrophic cancellation ?
        Lnn (ccm/sub (ccm/identity-matrix n)
                     (ccm/mmul Knm Hmn))
        P''nn (ccm/mmul Lnn P'nn)] ; New covariance

    {:xn1 x''nl, :Pnn P''nn}))
```

9.4.1 UNIT TEST

Let the measurement model be a cubic:

```
(defn Hmn-t [t]
  (ccm/matrix [( * t t t) (* t t) t 1]))
```

Ground truth state, constant with time in this unit test:

```
(def true-x
  (ccm/array [-5 -4 9 -3]))

(require '[clojure.core.matrix.random :as ccmr])
```

```
(defn fake [n]
  (let [times (range -2.0 2.0 (/ 2.0 n))
        Hmns (mapv Hmn-t times)
        true-zs (mapv #(ccm/mmul % true-x) Hmns)
        zmls (mapv #(ccm/add
                      % (ccm/array
                        [(ccmr/rand-gaussian)]))
                    true-zs)]
    {:times times, :Hmns Hmns, :true-zs true-zs, :zmls zmls}))
```

```
1188 (def test-data (fake 7))
```

```
1189   A state cluster is a vector of  $\mathbf{x}$  and  $\mathbf{P}$ :
```

```
1190 (def state-cluster-prior
1191   {:xn1 (ccm/array [[0.0] [0.0] [0.0] [0.0]])
1192    :Pnn (ccm/mul 1000.0 (ccm/identity-matrix 4))})
```

```
1193   An obn-cluster is a vector of  $\mathbf{z}$ ,  $\mathbf{H}$ ,  $\mathbf{A}$ ,  $\mathbf{Q}$ , and  $\mathbf{R}$ . Obn is short for observation.
```

```
1194 (def obn-clusters
1195   (let [c (count (:times test-data))]
1196     (mapv (fn [zml Hmn Ann Qnn Rmm]
1197             {:zml zml, :Hmn Hmn, :Ann Ann, :Qnn Qnn, :Rmm Rmm})
1198           (:zmls test-data)
1199           (:Hmns test-data)
1200           (repeat c (ccm/identity-matrix 4))
1201           (repeat c (ccm/zero-matrix 4 4))
1202           (repeat c (ccm/identity-matrix 1))
1203           )))
```

```
1204 (clojure.pprint/pprint (reduce kalman-update state-cluster-prior obn-clusters))
```

```
1205 {:xn1 #vectorz/matrix [[-4.900886700347095],
1206 [-4.1922940308673535],
1207 [8.804508118620738],
1208 [-2.1922826490975567]],
1209 :Pnn
1210 #vectorz/matrix [[0.03208215055213958,-5.478256737134757E-15,-0.0874691388122202,-8.77076
1211 [-2.3568386825489895E-15,0.03637145347999561,-5.2632377622874316E-14,-0.05541947257604415],
1212 [-0.08746913881223455,-2.570860191397628E-14,0.2822249372573019,-1.1334683192032458E-14],
1213 [4.6455894686658894E-15,-0.05541947257607027,-6.734196533741965E-15,0.15110531309503664]]]}
```

```
1214   Notice how close the estimate  $\mathbf{x}_{n \times 1}$  is to the ground truth,  $[-5, -4, 9, -3]$  for  $\mathbf{x}$ . A chi-squared test would
1215 be appropriate to complete the verification (TODO).
```

1216 9.5 DEFN MAKE-KALMAN-MAPPER

```
1217 Just as we did before, we can convert a foldable into a mappable transducer and bang on an asynchronous
1218 stream of data. This only needs error handling to be deployable at scale. Not to minimize error handling:
1219 it's a big but separable engineering task.
```

```
1220 (do (defn make-kalman-mapper [{:keys [xn1 Pnn]})
1221     ;; let-over-lambda (LOL); here are the Bayesian priors
1222     (let [estimate-and-covariance (atom {:xn1 xn1, ;; prior-estimate
1223                                           :Pnn Pnn, ;; prior-covariance
1224                                           })]
1225       ;; here is the mapper (mappable)
1226       (fn [{:keys [zml Hmn Ann Qnn Rmm]}]
1227         (let [{:xn1 :xn1, Pnn :Pnn} @estimate-and-covariance]
1228           (let [;; out-dented so we don't go crazy reading it
1229                 x'nl (ccm/mmul Ann xn1) ; Predict state
1230                 P'nn (ccm/add
1231                       Qnn (similarity-transform Ann Pnn)) ; Predict covariance
1232                 Dmm (ccm/add
1233                       Rmm (similarity-transform Hmn P'nn)) ; Gain precursor
1234                 DTmm (ccm/transpose Dmm) ; Support for "solve"
```

```

1235 HP'Tmn (ccm/mmul Hmn (ccm/transpose P'nn)) ; Support for "solve"
1236 ; Eqn 3 of http://vixra.org/abs/1606.0328
1237 KTmn (solve-matrix DTmm HP'Tmn)
1238 Knm (ccm/transpose KTmn) ; Kalman gain
1239 ; innovation = predicted obn residual
1240 rml (ccm/sub zml (ccm/mmul Hmn x'n1))
1241 x''n1 (ccm/add x'n1 (ccm/mmul Knm rml)) ; final corrected estimate
1242 n (ccm/dimension-count xn1 0)
1243 ; new covariance ? catastrophic cancellation ?
1244 Lnn (ccm/sub (ccm/identity-matrix n)
1245 (ccm/mmul Knm Hmn))
1246 P''nn (ccm/mmul Lnn P'nn)]
1247 (swap! estimate-and-covariance conj
1248 [:xn1 x''n1]
1249 [:Pnn P''nn]) ) )
1250 @estimate-and-covariance ) )
1251
1252 ;; The following line maps over a fixed sequence in memory
1253 #_(clojure.pprint/pprint (last
1254 (map (make-kalman-mapper state-cluster-prior)
1255 obn-clusters)))
1256
1257 #_(async-randomized-scan obn-clusters
1258 (make-kalman-mapper state-cluster-prior)
1259 clojure.pprint/pprint)
1260
1261 (let [accumulator (make-sow-reap)]
1262 (async-randomized-scan obn-clusters
1263 (make-kalman-mapper state-cluster-prior)
1264 (accumulator ::sow))
1265 (last (accumulator ::reap))) )
1266 '(:xn1 #vectorz/matrix ((-4.900886700346977)
1267 (-4.192294030867509)
1268 (8.8045081186206)
1269 (-2.192282649097485)) :Pnn #vectorz/matrix ((0.03208215055212902 -5.7829175476031836E-15 -
1270 (3.0805239815186874E-14 0.0363714534799503 -1.0213615435177015E-13 -0.055419472576001716)
1271 (-0.08746913881225812 6.337638747133667E-14 0.28222493725721515 -3.8163916471489756E-16)
1272 (3.456436525883788E-15 -0.055419472576089764 3.8615811936981714E-14 0.15110531309502237)))

```

10 VISUALIZATION

10.1 INCANTER

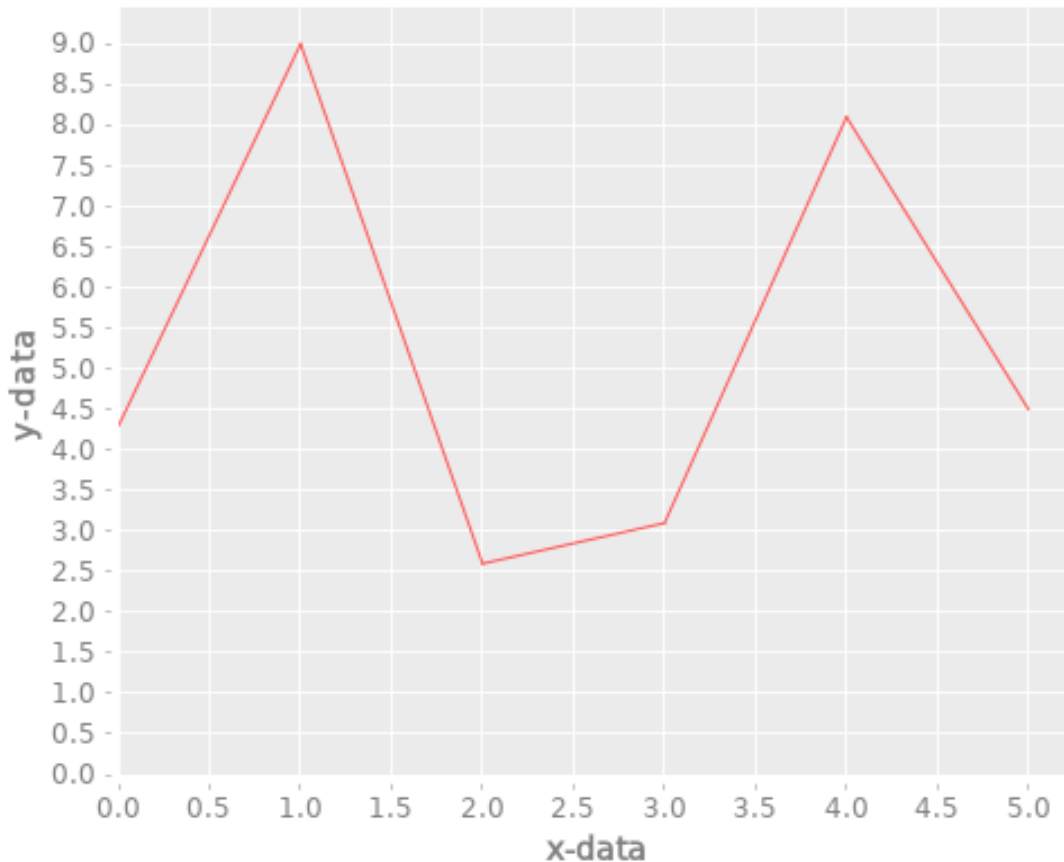
Hot-loading seems hopelessly broken.

```

1276 (cljr-add-project-dependency)

1277 (use '(incanter core charts pdf))
1278 ;;; Create the x and y data:
1279 (def x-data [0.0 1.0 2.0 3.0 4.0 5.0])
1280 (def y-data [4.3 9.0 2.6 3.1 8.1 4.5])
1281 (def xy-line (xy-plot x-data y-data))
1282 #_(view xy-line)
1283 (save-pdf xy-line "incanter-xy-line.pdf")
1284 (save xy-line "incanter-xy-line.png")

```



1285

1286 10.2 OZ

1287 From [https://github.com/metasoarous/oz/blob/master/examples/clojupyter-example.](https://github.com/metasoarous/oz/blob/master/examples/clojupyter-example.ipynb)
 1288 ipynb

```
1289 (require '[clojupyter.misc.helper :as helper])
1290 (helper/add-dependencies '[metasoarous/oz "1.6.0-alpha2"])
1291 (require '[oz.notebook.clojupyter :as oz])
```

1292 10.2.1 DEFN PLAY DATA

```
1293 (do (defn play-data [& names]
1294     (for [n names
1295           i (range 20)]
1296       {:time i :item n :quantity (+ (Math/pow (* i (count n)) 0.8) (rand-int (count n)))})
1297
1298     (def stacked-bar
1299       {:data {:values (play-data "munchkin" "witch" "dog" "lion" "tiger" "bear")}
1300        :mark "bar"
1301        :encoding {:x {:field "time"}
1302                   :y {:aggregate "sum"
1303                        :field "quantity"
1304                        :type "quantitative"}
1305                   :color {:field "item"}}})
1306     (oz/view! stacked-bar) )
```

```
1307 #object[oz.notebook.clojupyter$view_BANG_$reify__21517 0x6473a298 "oz.notebook.clojupyter$
1308 ;; Create spec, then visualize
1309 (def spec
1310   {:data {:url "https://gist.githubusercontent.com/metasoarous/4e6f781d353322a44b9cd3e45970
1311     :mark "point"
1312     :encoding {
1313       :x {:field "Horsepower", :type "quantitative"}
1314       :y {:field "Miles_per_Gallon", :type "quantitative"}
1315       :color {:field "Origin", :type "nominal"}}})
1316   (oz/view! spec)

1317 #'composable-statistics.core/spec#object[oz.notebook.clojupyter$view_BANG_$reify__21517 0x
1318 (oz/view!
1319   [:div
1320     [:h1 "A little hiccup example"]
1321     [:p "Try drinking a glass of water with your head upside down"]
1322     [:div {:style {:display "flex" :flex-direction "row"}}
1323       [:vega-lite spec]
1324       [:vega-lite stacked-bar]]])

1325 #object[oz.notebook.clojupyter$view_BANG_$reify__21517 0x7d4aec9e "oz.notebook.clojupyter$
```

11 GAUSSIAN PROCESSES

The Extended Kalman Filter above is a generalization of linear regression.

11.1 RECURRENT LINEAR REGRESSION

Emacs 26.2 of 2019-04-12, org version: 9.2.2