PRINCETON UNIVERSITY

ORF 474: High Frequency Trading
Spring 2020
Robert Almgren

Lecture 6b

March 11, 2020

Today

- Futures products and data
- Kdb+ database system

Futures data and Kdb+

- Futures products
  rich product structure
- Futures data
  cleaner market data
  sequence numbers, timestamps, aggressor flag
- Database software
  Kdb+/Q, industry standard for large databases
  our course server configuration
  rkdb connector between R and Kdb+

Futures

- Stocks and bonds are *securities*
  - When you buy a stock, you own a "thing"
    - margin rules and short positions make it a bit more complicated
  - centrally cleared (at least in US -- DTCC)
  - market fragmentation (TAQ dataset)
- Futures and options are derivatives
  - standardized version of bilateral contract
  - clear at exchange (monolithic model)
  - no fragmentation
  - rich variety of products

Forward contract

Wheat farmer

Wheat

Money

On some future date, when wheat is grown and needed for bread

Baker

Taking care to maintain these relationships may not result in the lowest wheat price imaginable but does produce a stronger, more resilient supply network. In 2008, for instance, unprocessed organic wheat prices shot from $8 per bushel, to $15, then $25, and even $50. Many users simply could not get the wheat they wanted in that environment because the farmers supplying those networks could make so much more money by disregarding existing contracts and selling to the highest bidder. Our farmers, on the other hand, held the line on the prices that they had committed themselves to sell at and filled their contracts at $25, despite considerable temptation to seek even higher prices.

CME (CBOT) rulebook

A *futures* is a contract to buy or sell in the future with specified terms

Chapter 14
Wheat Futures

14100. SCOPE OF CHAPTER

This chapter is limited in application to futures trading of wheat. The procedures for trading, clearing, inspection, delivery and settlement, and any other matters not specifically covered herein or in Chapter 7 shall be governed by the general rules of the Exchange.

14101. CONTRACT SPECIFICATIONS

Each futures contract shall be for 5,000 bushels of No. 2 Soft Red Winter, No. 2 Hard Red Winter, No. 2 Dark Northern Spring, and No. 2 Northern Spring at par; and No. 1 Soft Red Winter, No. 1 Hard Red Winter, No. 1 Dark Northern Spring and No. 1 Northern Spring at 3 cents per bushel over contract price. Every delivery of wheat may be made up of the authorized grades for shipment from eligible regular facilities provided that no lot delivered shall contain less than 5,000 bushels of any one grade in any one facility.

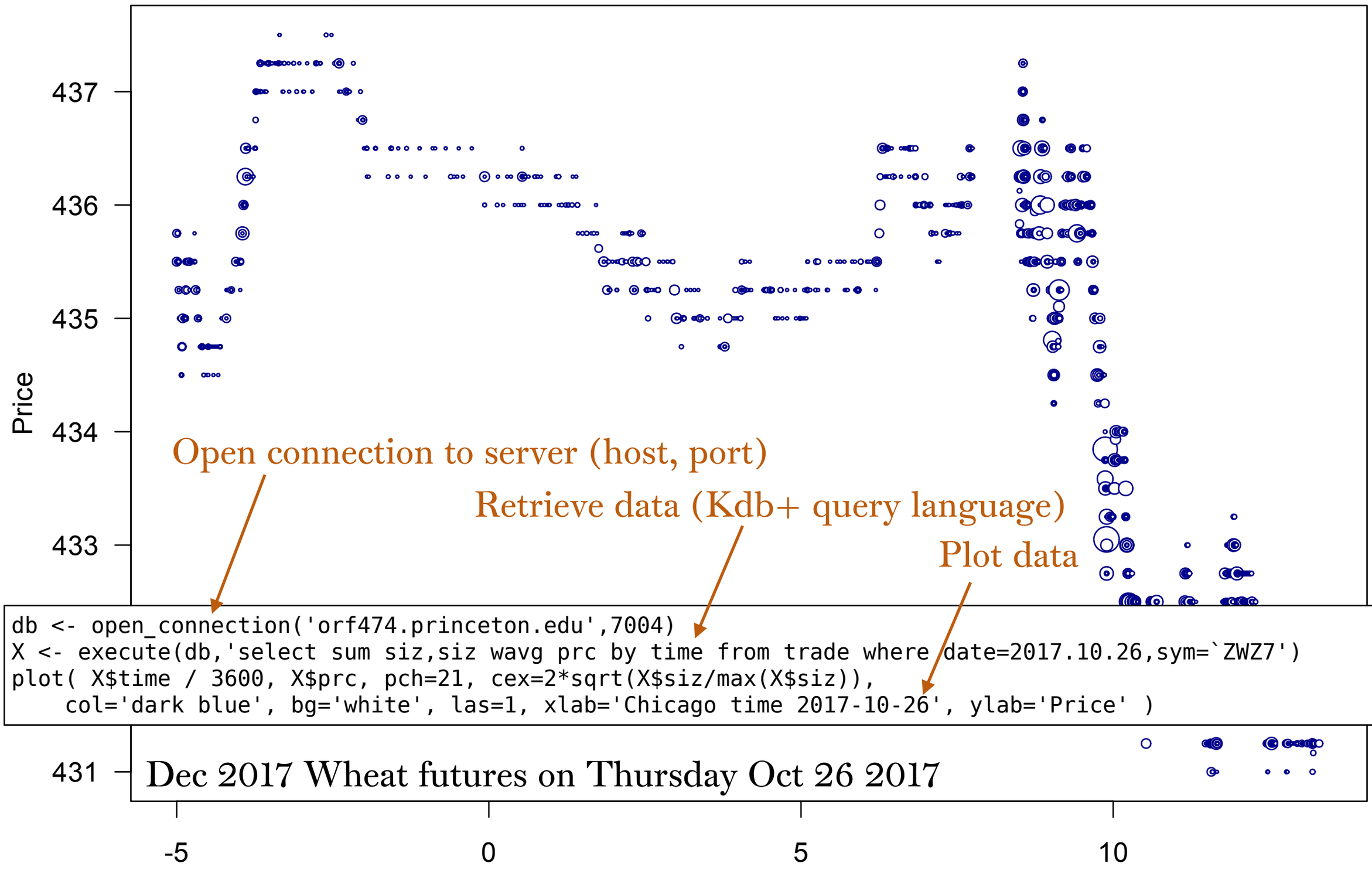
14104. GRADES / GRADE DIFFERENTIALS[1]

Upon written request by a taker of delivery at the time loading orders are submitted, a futures contract for the sale of wheat shall be performed on the basis of United States origin only.

WHEAT GRADE DIFFERENTIALS	
At 3¢ Premium	**At Contract Price**
No. 1 Soft Red Winter	No. 2 Soft Red Winter
No. 1 Hard Red Winter	No. 2 Hard Red Winter
No. 1 Dark Northern Spring	No. 2 Dark Northern Spring
No. 1 Northern Spring	No. 2 Northern Spring

Wheat which contains moisture in excess of 13.5% is not deliverable.

High-frequency trading of wheat futures



Types of futures contracts

- Agricultural

  wheat, corn, soybean, hog, cattle, milk, ...
- Energy

  crude oil, heating oil, gasoline, natural gas, ...
- Interest rates

  Treasury notes/bonds, pure rates (cash-settled), ...
- Equity index

  S&P 500, Nasdaq, ...
- Foreign exchange

  rates on USD vs foreign
- Precious metals

  gold, silver, ...

Settlement

- Physical settlement

  exchange money for object (wheat, etc)
- Cash settlement

  exchange cash based on numerical value

  equity index futures

  Eurodollars: settle based on LIBOR rate

- Mark to market:

  long/short positions exchange cash each day based on changes in contract price

- Very few futures contracts go to delivery

  mostly used by speculators

Futures exchanges worldwide

- CME Group (Chicago)
  - Chicago Mercantile Exchange (Ags, Rates)
  - Chicago Board of Trade (Treasuries, equity)
  - NYMEX (energy)
  - Comex (precious metals)
- Eurex (Frankfurt)-- interest rates and equity index
- LIFFE (now ICE Europe -- London) -- interest rates
- ICE (New York) -- energy, ag, equity
- CBOE -- VIX volatility
- Montreal -- interest rates, equity
- ASX (Sydney) -- interest rates, equity
- etc

Advantages of futures for microstructure study (compared to cash equities)

- Single exchange: no fragmentation
- Extremely liquid and actively traded

  price discovery often happens in futures
- Interesting range of product types

  interest rates, equities, etc
- Interesting collection of trading rules

  time priority, pro rata, etc
- Interesting relationships between different products

  cointegration, implied quoting, etc

Disadvantages of futures for microstructure study (compared to cash equities)

- More exotic products
  - harder to understand
  - less familiar
  - less central to society
- Smaller range of tick size
  - exchange adjusts tick for each product

Futures symbology

Futures contracts have

- "instrument" name (general class)
  e.g. "ES" for SP500 futures
- maturity month and year
  "Z" for December
  "7" for 2017
  "Z7" = December 2017
- combination is the "symbol"
  ESZ7 = SP500 Dec 2017

Root of symbol "instrument"

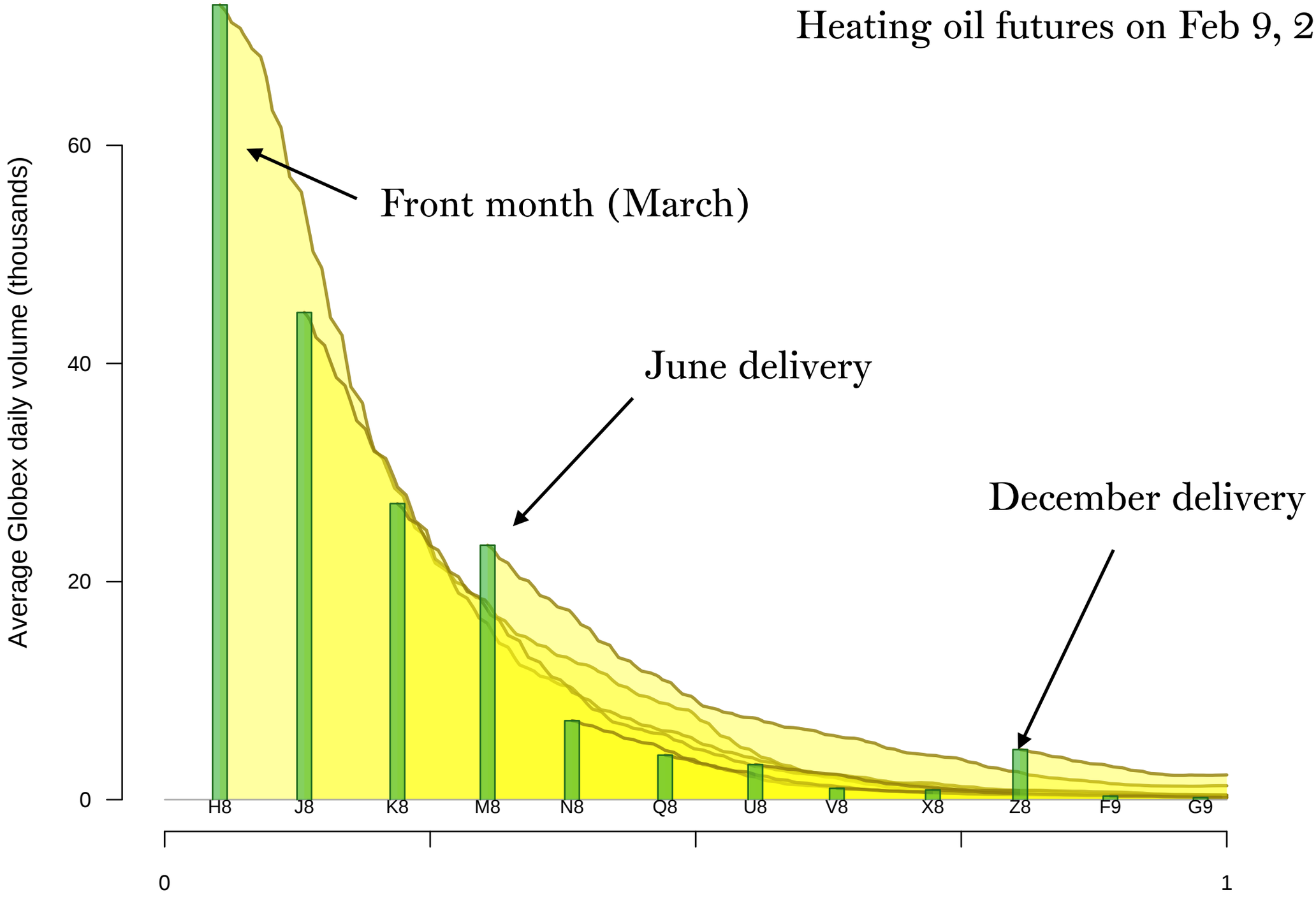
Product

inst	class	name	subclass	cur	notional	minpxincr	minpxincrval	dispfactor	matchalgo
ZC	AG	Corn	Grain And Oilseed	USD	50	0.25	12.5	1	K
KE	AG	KC Hard Red Winter (HRW) Wheat	Grain And Oilseed	USD	50	0.25	12.5	1	O
ZS	AG	Soybean	Grain And Oilseed	USD	50	0.25	12.5	1	K
ZM	AG	Soybean Meal	Grain And Oilseed	USD	10	1	10	0.1	K
ZL	AG	Soybean Oil	Grain And Oilseed	USD	6	1	6	0.01	K
ZW	AG	Wheat	Grain And Oilseed	USD	50	0.25	12.5	1	K
GF	AG	Feeder Cattle	Livestock	USD	0.5	25	12.5	0.001	K
HE	AG	Lean Hog	Livestock	USD	0.4	25	10	0.001	K
LE	AG	Live Cattle	Livestock	USD	0.4	25	10	0.001	K
CL	EN	Crude Oil	Crude Oil	USD	10	1	10	0.01	F
CLT	EN	TAS Crude Oil	Crude Oil	USD	10	1	10	0.01	F
NG	EN	Natural Gas (Henry Hub) Physical	Natural Gas	USD	10	1	10	0.001	F
NGT	EN	TAS Natural Gas (Henry Hub) Physical	Natural Gas	USD	10	1	10	0.001	F
HO	EN	NY Harbor ULSD	Refined Products	USD	4.2	1	4.2	0.0001	F
RB	EN	RBOB Gasoline Physical	Refined Products	USD	4.2	1	4.2	0.0001	F
HOT	EN	TAS NY Harbor ULSD	Refined Products	USD	4.2	1	4.2	0.0001	F
RBT	EN	TAS RBOB Gasoline Physical	Refined Products	USD	4.2	1	4.2	0.0001	F
RTY	EQ	R2000 Ind Mini		USD	0.5	10	5	0.01	F
NKD	EQ	Nikkei/USD	International Index	USD	5	5	25	1	F
NIY	EQ	Nikkei/Yen	International Index	JPY	500	5	2500	1	F
YM	EQ	E-mini Dow ($5)	US Index	USD	5	1	5	1	F
NQ	EQ	E-mini NASDAQ 100	US Index	USD	0.2	25	5	0.01	F
ES	EQ	E-mini S&P 500	US Index	USD	0.5	25	12.5	0.01	F
SMC	EQ	E-mini S&P 600 SmallCap	US Index	USD	1	10	10	0.01	F
EMD	EQ	E-mini S&P MidCap 400	US Index	USD	1	10	10	0.01	F
6M	FX	Mexican Peso	Emerging Market	USD	0.5	10	5	1e-06	F
6A	FX	Australian Dollar	Majors	USD	10	1	10	0.0001	F
6B	FX	British Pound	Majors	USD	6.25	1	6.25	0.0001	F
6C	FX	Canadian Dollar	Majors	USD	10	0.5	5	0.0001	F
6E	FX	Euro FX	Majors	USD	12.5	0.5	6.25	0.0001	F
6J	FX	Japanese Yen	Majors	USD	12.5	0.5	6.25	1e-06	F
6N	FX	New Zealand Dollar	Majors	USD	10	1	10	0.0001	F
6S	FX	Swiss Franc	Majors	USD	12.5	1	12.5	0.0001	F

...

Multiple expiration dates for each product

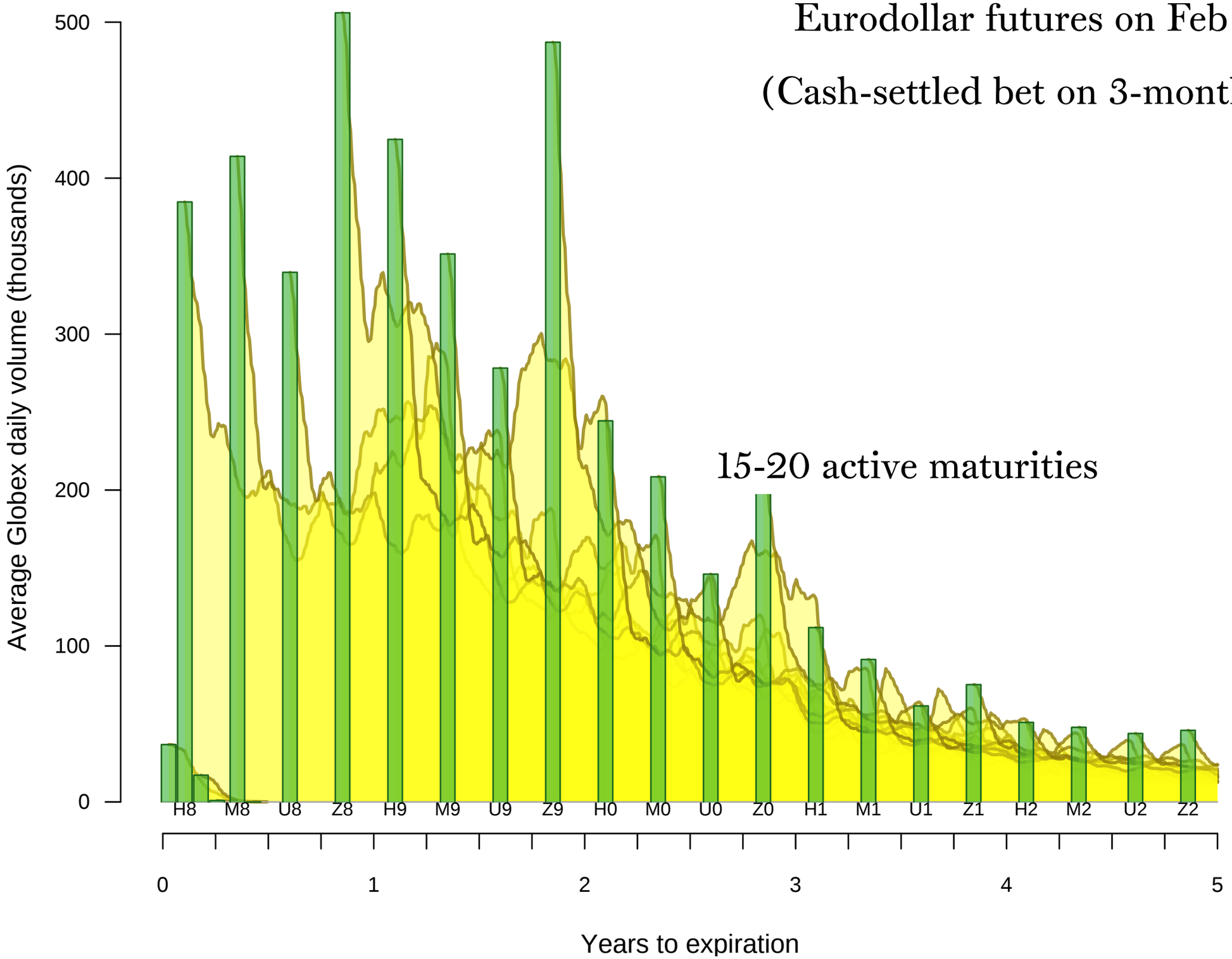
Heating oil futures on Feb 9, 2018



Broad range of active maturities

Eurodollar futures on Feb 9, 2018

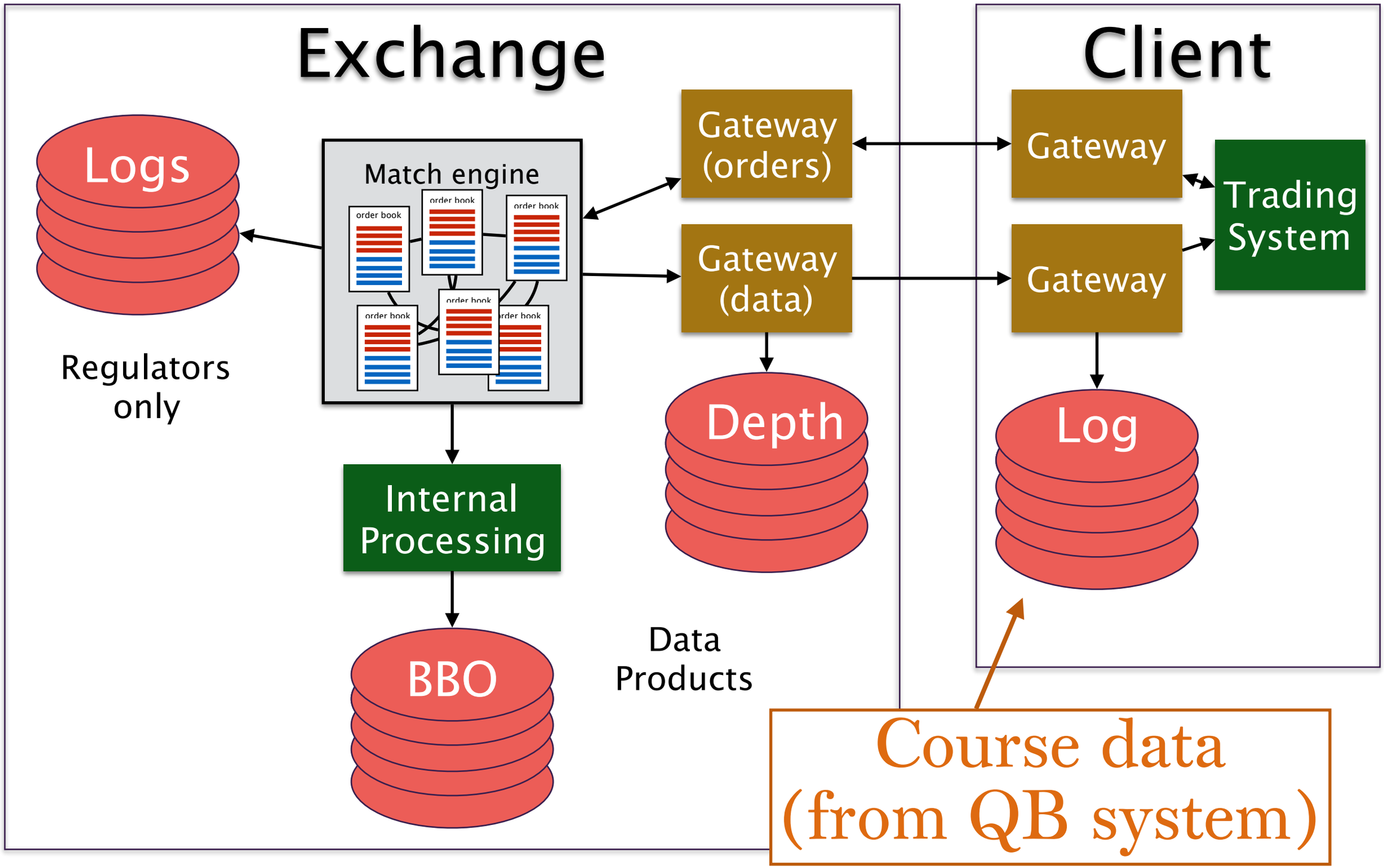
(Cash-settled bet on 3-month LIBOR)



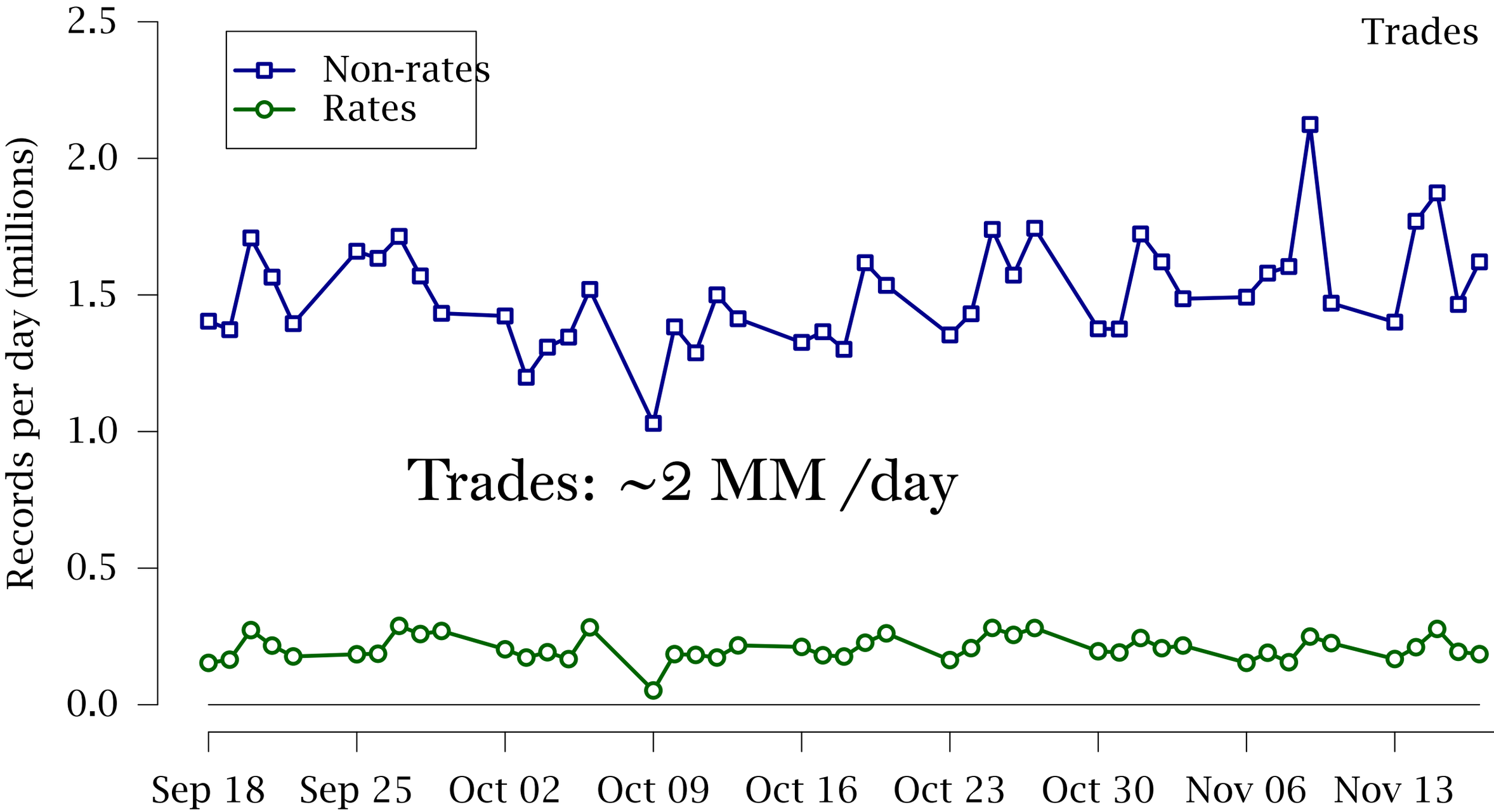
Database content

- All CME tick data for 9 weeks
  Monday Sept 18 through Friday Nov 17, 2017
- All products (rates, energy, equity index, etc)
  Interest rates / non-rates on different services
- All outrights and spreads
- Very precise and reliable time stamps

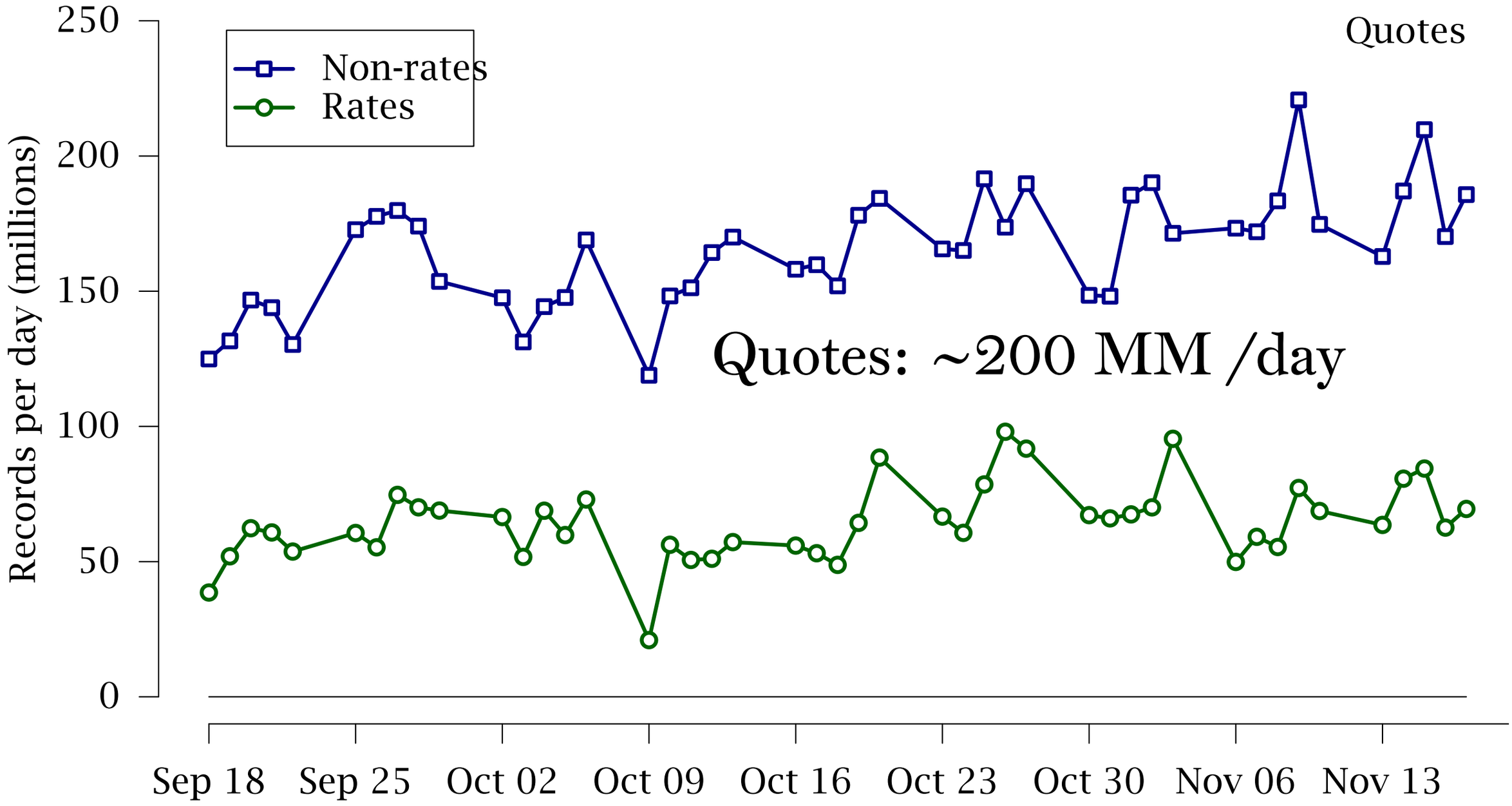
Data recording from futures exchange



Number of trade records per day



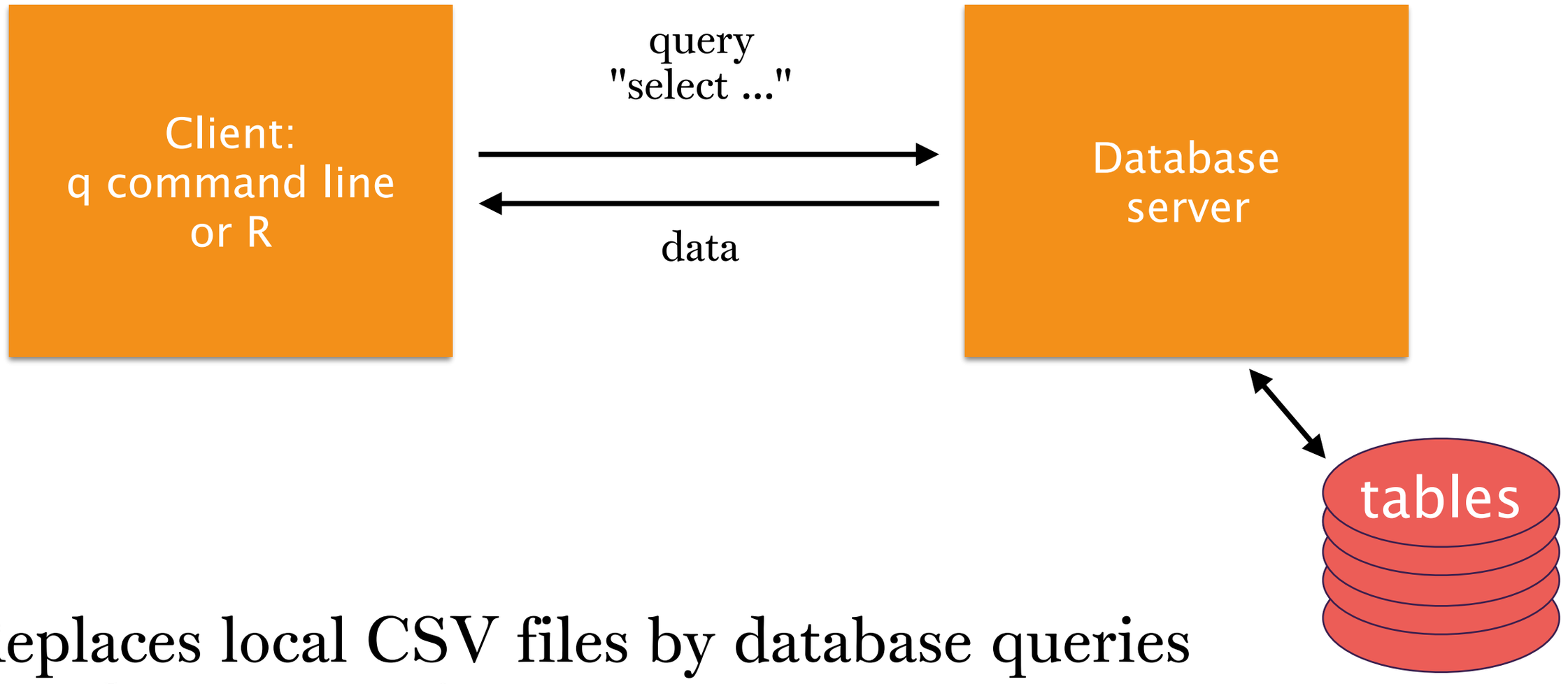
Number of quote records per day



How to handle large data

- Web page and CSV file download (WRDS TAQ)
  - Simple, low barrier to entry
  - cumbersome for large variety of data
- Database software
  - designed to manage data
  - client-server protocol
  - data directly into application (R)
  - somewhat standardized syntax: SQL

Client-server database



- Replaces local CSV files by database queries
- Can do computations on remote server

Standard software on Wall Street

- Kdb+ or Q (http://www.kx.com)
- Descendant of APL
- Designed for large time series databases
- 64-bit client free to download

  Mac Catalina does not support 32-bit binaries
- Programming language as well as database

Q references

- Kx main code page
  `https://code.kx.com/q/`
- "Q for Mortals"
  `http://code.kx.com/q4m3/`
- R connector
  `https://github.com/KxSystems/rkdb`
- Command line client
  `https://kx.com/download/`

Our server

- Host: `orf474.princeton.edu`
  accessible only from within Princeton
- Interest rates on ports `6000,...,6009`
- Non-rates on ports `7000,...,7009`
- End-of-day data on ports `8000, ...,8009`
- 1-minute samples of Crude Oil on `9000, ...,9009`

  Identical data on each port within each category

Q basics

- tables
- queries
- handles
- variables and data types
  - assignment
- arithmetic
  - division, right-to-left precedence
- lists
- functions
- joins

Queries

set console window size

`hopen host:port` returns a "handle"

`h"command"` runs the command on the remote server

Query syntax somewhat like SQL

```
q)\c 50 200

q)h:hopen `$":orf474.princeton.edu:6007"

q)h"tables[]"
`s#`instinfo`matchalgos`quote`trade`trdords

q)h"select from instinfo where class=`IR"
inst class name                                        subclass          cur notional minpxincr minpxincrva..
-------------------------------------------------------------------------------------------------------------..
B1U  IR    30 Year USD Deliverable Interest Rate Swap  Deliverable Swaps USD 1000     0.03125   31.25       ..
F1U  IR    5 Year USD Deliverable Interest Rate Swap   Deliverable Swaps USD 1000     0.015625  15.625      ..
N1U  IR    10 Year USD Deliverable Interest Rate Swap  Deliverable Swaps USD 1000     0.015625  15.625      ..
T1U  IR    2 Year USD Deliverable Interest Rate Swap   Deliverable Swaps USD 1000     0.0078125 7.8125      ..
GE   IR    Eurodollar                                  Stirs             USD 25       0.5       12.5        ..
ZQ   IR    30 Day Federal Funds                        Stirs             USD 41.67    0.5       20.835      ..
TN   IR    Ultra 10 Yr Notes                           US Treasury       USD 1000     0.015625  15.625      ..
UB   IR    Ultra U.S. Treasury Bond                    US Treasury       USD 1000     0.03125   31.25       ..
Z3N  IR    3-Year Note                                 US Treasury       USD 2000     0.0078125 15.625      ..
ZB   IR    U.S. Treasury Bond                          US Treasury       USD 1000     0.03125   31.25       ..
ZBX  IR    3:2 U.S. Treasury Bond                      US Treasury       USD 1000     0.03125   31.25       ..
ZF   IR    5 Yr Note                                   US Treasury       USD 1000     0.0078125 7.8125      ..
ZN   IR    10 Yr Note                                  US Treasury       USD 1000     0.015625  15.625      ..
ZT   IR    2 Yr Note                                   US Treasury       USD 2000     0.0078125 15.625      ..
```

Trade and quote tables

Sequence number
consistent between
trades and quotes

time =
exchange time stamp
nanosecond precision
same for all events resulting
from same inbound order

```
q)h"10 # select from trade where date=2017.10.26,sym=`ZNZ7"
date       sym  seq     time                         prc       siz aggr
-----------------------------------------------------------------------
2017.10.26 ZNZ7 6395193 -0D06:59:59.998940858 124.546875 331
2017.10.26 ZNZ7 6395201 -0D06:59:59.998940858 124.5625   5
2017.10.26 ZNZ7 6395228 -0D06:59:59.986967891 124.5625   5   B
2017.10.26 ZNZ7 6395236 -0D06:59:59.983814135 124.5625   5   B
2017.10.26 ZNZ7 6395239 -0D06:59:59.982338831 124.546875 6   B
2017.10.26 ZNZ7 6395243 -0D06:59:59.981244855 124.5468   6   B
2017.10.26 ZNZ7 6395257 -0D06:59:59.955264077 124.5468   6   B
2017.10.26 ZNZ7 6395274 -0D06:59:59.791841209 124.5468   6   B
2017.10.26 ZNZ7 6395284 -0D06:59:59.465341799 124.5625   5   B
2017.10.26 ZNZ7 6395317 -0D06:59:59.386660911 124.5625   5   B
```

Always put date
as first condition
in where clause

```
q)h"10 # select from quote where date=2017.10.26,sym=`ZNZ7"
date       sym  seq     time                         bid       ask       bsiz asiz..
-------------------------------------------------------------------------------..
2017.10.26 ZNZ7 6394713 -0D07:12:05.006493889 124.546875             461       ..
2017.10.26 ZNZ7 6394713 -0D07:12:05.006493889 124.546875 124.546875 461  225  ..
2017.10.26 ZNZ7 6394781 -0D07:06:41.647755697 124.546875 124.546875 499  225  ..
2017.10.26 ZNZ7 6394787 -0D07:06:07.017817005 124.546875 124.546875 499  227  ..
2017.10.26 ZNZ7 6394812 -0D07:05:12.603337735 124.546875 124.546875 499  207  ..
2017.10.26 ZNZ7 6394824 -0D07:05:10.002855099 124.546875 124.546875 505  207  ..
2017.10.26 ZNZ7 6394840 -0D07:04:59.992506177 124.546875 124.546875 506  207  ..
2017.10.26 ZNZ7 6394842 -0D07:04:59.891321331 124.546875 124.546875 508  207  ..
2017.10.26 ZNZ7 6394844 -0D07:04:59.890905067 124.546875 124.546875 508  206  ..
2017.10.26 ZNZ7 6394859 -0D07:04:59.300948099 124.546875 124.546875 508  140  ..
```

Time stamp field

`time` = Fix Tag 60
Same value for all market data updates
resulting from one inbound message
(converted to Chicago time in our database)

Message Layout

35 = X Message Body

Tag	Name	FIX Type	Enumeration	Req	Description
60	TransactTime	UTCTimestamp		Y	Start of event processing time (UTC). UTC Timestamps are sent in number of nanoseconds since Unix epoch synced to a master clock to microsecond accuracy.

`seq` = Fix Tag 83
Consecutive across trades and quotes for each symbol

→83	RptSeq	Int		C	MD Entry sequence number per instrument update. Reset weekly.

`https://www.cmegroup.com/confluence/display/EPICSANDBOX/MDP+3.0+-+Market+Data+Incremental+Refresh`

timestamp field

All these trades result
from one incoming order

```
q)h"select from trade where date=2017.10.26,time=0D07:03:48.305288099"
date       sym       seq     time                          prc    siz aggr
-----------------------------------------------------------------------
2017.10.26 GEH2      7333159 0D07:03:48.305288099 9749.5 7
2017.10.26 GEM0      9410439 0D07:03:48.305288099 9775.5 23
2017.10.26 GEM0-GEM1 1523770 0D07:03:48.305288099 15     23
2017.10.26 GEM1      8568081 0D07:03:48.305288099 9760.5 188 S
2017.10.26 GEM1-GEH2 1205433 0D07:03:48.305288099 11     7
2017.10.26 GEM1-GEU1 1313338 0D07:03:48.305288099 3.5    21
2017.10.26 GEM1-GEZ1 1213975 0D07:03:48.305288099 8      4
2017.10.26 GEU1      8715117 0D07:03:48.305288099 9757   21
2017.10.26 GEU9      8247900 0D07:03:48.305288099 9788   9
2017.10.26 GEU9-GEM1 1132288 0D07:03:48.305288099 27.5   9
2017.10.26 GEZ0      9673866 0D07:03:48.305288099 9767.5 4
2017.10.26 GEZ0-GEM1 1346543 0D07:03:48.305288099 7      4
2017.10.26 GEZ1      8206682 0D07:03:48.305288099 9752.5 4
```

Joining trades and quotes

This function is defined already on the server

Left and right sequence numbers

```
q)h"tqmergeS[2017.10.26;`ZNZ7;8001062;8001083]"
date       sym  seq     time                 bsiz bid       siz prc        aggr ask        asiz
-------------------------------------------------------------------------------------------
2017.10.26 ZNZ7 8001062 0D09:05:12.137057453 578  124.53125                    124.546875 2165
2017.10.26 ZNZ7 8001063 0D09:05:12.137394473 578  124.53125                    124.546875 2179
2017.10.26 ZNZ7 8001064 0D09:05:12.315486879 578  124.53125                    124.546875 2174
2017.10.26 ZNZ7 8001066 0D09:05:12.690400123 578  124.53125                    124.546875 2163
2017.10.26 ZNZ7 8001069 0D09:05:13.376149393                 90  124.53125  S
2017.10.26 ZNZ7 8001071 0D09:05:13.376149393 488  124.53125                    124.546875 2163
2017.10.26 ZNZ7 8001072 0D09:05:13.376513257 487  124.53125                    124.546875 2163
2017.10.26 ZNZ7 8001073 0D09:05:13.376531117 487  124.53125                    124.546875 2164
2017.10.26 ZNZ7 8001074 0D09:05:13.384631865 487  124.53125                    124.546875 2174
2017.10.26 ZNZ7 8001075 0D09:05:13.384847679 477  124.53125                    124.546875 2174
2017.10.26 ZNZ7 8001078 0D09:05:13.530712775 477  124.53125                    124.546875 2171
2017.10.26 ZNZ7 8001080 0D09:05:13.565074593                 11  124.546875 B
2017.10.26 ZNZ7 8001082 0D09:05:13.565074593 477  124.53125                    124.546875 2160
2017.10.26 ZNZ7 8001083 0D09:05:13.565378257 478  124.53125                    124.546875 2160
```

query syntax (like SQL but not exactly)

Columns or functions of columns

`trade` or `quote`

Conditions on date, time, symbol, etc (comma for "and")

```
select ... from ... where ...
```

Grouping variable

```
select ... by ... from ... where ...
```

joins

first 10 trades after 10:00:00

```
q)h"10 # select seq,time,prc,siz,aggr from trade where date=2017.10.26,sym=`ZNZ7,time>10:00:00"
seq     time                 prc        siz aggr
-----------------------------------------------
8195366 0D10:00:00.012026921 124.484375 946 B
8195895 0D10:00:00.989004977 124.46875  1   S
8196087 0D10:00:03.315642729 124.46875  2   S
8196095 0D10:00:03.737349063 124.46875  90  S
8196118 0D10:00:04.225435593 124.484375 339 B
8196291 0D10:00:05.848141313 124.484375 2   B
8196403 0D10:00:08.670303591 124.484375 50  B
8196473 0D10:00:10.345328539 124.46875  2   S
8196530 0D10:00:11.179054915 124.484375 10  B
8196538 0D10:00:11.333296783 124.484375 12  B
```

Determine quote bid, ask, bsiz, asiz, time for each trade

`aj` = "asof join"

```
q)h"aj[ `seq;
    10 # select seq,time,prc,siz,aggr from trade where date=2017.10.26,sym=`ZNZ7,time>10:00:00;
    select qtime:last time,last bid,last ask,last bsiz,last asiz by seq from quote where date=2017.10.26,sym=`ZNZ7 ]"
seq     time                 prc        siz aggr qtime                bid       ask       bsiz asiz
-------------------------------------------------------------------------------------------------
8195366 0D10:00:00.012026921 124.484375 946 B    0D09:59:58.713630557 124.46875 124.484375 1138 2672
8195895 0D10:00:00.989004977 124.46875  1   S    0D10:00:00.926214307 124.46875 124.484375 1739 1386
8196087 0D10:00:03.315642729 124.46875  2   S    0D10:00:03.009595715 124.46875 124.484375 1867 1337
8196095 0D10:00:03.737349063 124.46875  90  S    0D10:00:03.324023045 124.46875 124.484375 1865 1337
8196118 0D10:00:04.225435593 124.484375 339 B    0D10:00:04.090634599 124.46875 124.484375 1848 1346
8196291 0D10:00:05.848141313 124.484375 2   B    0D10:00:05.747546181 124.46875 124.484375 1953 1068
8196403 0D10:00:08.670303591 124.484375 50  B    0D10:00:08.572888441 124.46875 124.484375 2078 1086
8196473 0D10:00:10.345328539 124.46875  2   S    0D10:00:10.029739343 124.46875 124.484375 2190 1043
8196530 0D10:00:11.179054915 124.484375 10  B    0D10:00:11.068705335 124.46875 124.484375 2197 977
8196538 0D10:00:11.333296783 124.484375 12  B    0D10:00:11.198655673 124.46875 124.484375 2198 950
```

Trades relative to bid and ask

first row has bid=ask

```
q)t:h"aj[ `seq;
   select seq,time,prc,siz,aggr from trade where date=2017.10.26,sym=`ZNZ7;
   select last bid,last ask,last bsiz,last asiz by seq from quote where date=2017.10.26,sym=`ZNZ7 ]"
q)5 # t
seq     time                    prc        siz aggr bid        ask        bsiz asiz
-----------------------------------------------------------------------------------
6395193 -0D06:59:59.998940858 124.546875 331      124.546875 124.546875 914  331
6395201 -0D06:59:59.998940858 124.5625   5        124.546875 124.5625   583  246
6395228 -0D06:59:59.986967891 124.5625   5   B    124.546875 124.5625   588  268
6395236 -0D06:59:59.983814135 124.5625   5   B    124.546875 124.5625   588  263
6395239 -0D06:59:59.982338831 124.546875 6   S    124.546875 124.5625   588  258

q)update npct:100*n%count t from select n:count i by s1:signum prc-bid,s2:signum prc-ask from t
s1 s2| n     npct
-----| ----------------------
-1 -1| 23    0.0413386534383
0  -1| 28303 50.869909055
0  0 | 1     0.00179733275819
1  -1| 103   0.185125274093
1  0 | 27181 48.8533017003
1  1 | 27    0.048527984471
```

prc>bid, prc>ask

prc=bid, prc<ask

prc = bid = ask

prc>bid, prc<ask

prc>bid, prc=ask

prc<bid, prc<ask

Connecting R to Kdb+

https://github.com/KxSystems/rkdb

```
install.packages('devtools')
library(devtools)
devtools::install_github('kxsystems/rkdb')
library(rkdb)
```

Open connection to server by host and port
Returned object is a "database handle"

```
> db <- open_connection('orf474.princeton.edu',7004)
```

Execute queries using that handle

```
> execute(db,'5 # select from trade where date=2017.10.26,sym=`ZWZ7')
        date  sym     seq             time    prc siz aggr
1 2017-10-26 ZWZ7 1378977 -17999998940218 435.50  21
2 2017-10-26 ZWZ7 1378991 -17999996243947 435.75   1
3 2017-10-26 ZWZ7 1379001 -17999996113875 435.75  16    B
4 2017-10-26 ZWZ7 1379021 -17998997128181 435.50   2    S
5 2017-10-26 ZWZ7 1379033 -17998982597985 435.50   1    S
```

Data type mapping

```
q)h"meta trade"
c    | t f a
-----| -----
date | d
sym  | s   p
seq  | i
time | n
prc  | f
siz  | i
aggr | s
```

```
> X <- execute(db,'5 # select from trade where date=2017.10.26,sym=`ZWZ7')
> colshow(X)
[1]          date Date
[1]           sym character
[1]           seq integer
[1]          time integer64
[1]           prc numeric
[1]           siz integer
[1]          aggr character
```

Nanosecond time types are mapped to R integer64 (nanoseconds since midnight)

```
> colshow
function (X)
{
    for (x in names(X)) print(paste(sprintf("%12s", x), paste(class(X[[x]]),
        collapse = ",")), quote = FALSE)
}
```

Useful tables on server

- `instinfo`: information about each instrument
- `sym2inst`: mapping from symbol to instrument
- `sym2type`: mapping from symbol to type
- `sym2expir`: mapping from symbol to expiration

  not guaranteed to always be correct
- `insts`: 30 most important symbols

Root of symbol "instrument"

inst	class	name	subclass	cur	notional	minpxincr	minpxincrval	dispfactor	matchalgo
KE	AG	KC Hard Red Winter (HRW) Wheat	Grain And Oilseed	USD	50	0.25	12.5	1	O
ZC	AG	Corn	Grain And Oilseed	USD	50	0.25	12.5	1	K
ZL	AG	Soybean Oil							
ZM	AG	Soybean Meal							
ZS	AG	Soybean							
ZW	AG	Wheat							
GF	AG	Feeder Cattle							
HE	AG	Lean Hog							
LE	AG	Live Cattle							
CL	EN	Crude Oil							
NG	EN	Natural Gas (Henry Hub) Physical	Natural Gas	USD	10	1	10	0.001	F
HO	EN	NY Harbor ULSD	Refined Products	USD	4.2	1	4.2	0.0001	F
RB	EN	RBOB Gasoline Physical	Refined Products	USD	4.2	1	4.2	0.0001	F
RTY	EQ	R2000 Ind Mini							
NIY	EQ	Nikkei/Yen							
NKD	EQ	Nikkei/USD							
EMD	EQ	E-mini S&P MidCap 400							
ES	EQ	E-mini S&P 500	US Index	USD	0.5	25	12.5	0.01	F
NQ	EQ	E-mini NASDAQ 100	US Index	USD	0.2	25	5	0.01	F
SMC	EQ	E-mini S&P 600 SmallCap	US Index	USD	1	10	10	0.01	F
YM	EQ	E-mini Dow ($5)	US I						
6M	FX	Mexican Peso	Eme						
6A	FX	Australian Dollar	Maj						
6B	FX	British Pound	Maj						
6C	FX	Canadian Dollar	Maj						
6E	FX	Euro FX	Maj						
6J	FX	Japanese Yen	Majors	USD	12.5	0.5	6.25	1e-06	F
6N	FX	New Zealand Dollar	Majors	USD	10	1	10	0.0001	F
6S	FX	Swiss Franc	Majors	USD	12.5	1	12.5	0.0001	F
B1U	IR	30 Year USD Deliverable Interest Rate Swap	Deliverable Swaps	USD	1000	0.03125	31.25	1	K
F1U	IR	5 Year USD Deliverable Interest Rate Swap	Deliverable Swaps	USD	1000	0.015625	15.625	1	K
N1U	IR	10 Year USD Deliverable Interest Rate Swap	Deliverable Swaps	USD	1000	0.015625	15.625	1	K
T1U	IR	2 Year USD Deliverable Interest Rate Swap	Deliverable Swaps	USD	1000	0.0078125	7.8125	1	K
GE	IR	Eurodollar	Stirs	USD	25	0.5	12.5	0.01	A
ZQ	IR	30 Day Federal Funds	Stirs	USD	41.67	0.5	20.835	0.01	K

Terminology for CME Globex trading system:

`inst` = generic name for product

`symbol` = `inst` + month code + last digit of year

F	G	**H**	J	K	**M**	N	Q	**U**	V	X	**Z**
Jan	Feb	**Mar**	Apr	May	**Jun**	Jul	Aug	**Sep**	Oct	Nov	**Dec**

`inst` = ES = SP500 futures

`symbol` = ESZ7 = Dec 2017 expiration

What Crude Oil products trade?

Port 7000-7009 for non-interest rate

Always put date condition first

Select products in the "CL" family using `sym2inst`

```
q)h:hopen `:orf474:7004

q)h"select v:sum siz by sym from trade where date=2017.10.24,sym2inst[sym]=`CL"
sym           | v
--------------| ----
CL:BF F8-G8-H8| 174
CL:BF F8-J8-N8| 1
CL:BF G8-H8-J8| 111
CL:BF H8-J8-K8| 85
CL:BF H8-M8-U8| 4
CL:BF J8-K8-M8| 15
CL:BF J9-K9-M9| 6
CL:BF K8-M8-N8| 739
CL:BF K9-M9-N9| 320
CL:BF M8-U8-Z8| 5
CL:BF M8-Z8-M9| 53
CL:BF M9-Z9-M0| 76
CL:BF N8-Q8-U8| 20
..
```

These are strange symbols! "Butterflies"

Classify by contract type using `sym2type`

```
q)h"select n:count distinct sym by typ:sym2type[sym] from trade where date=2017.10.24,sym2inst[sym]=`CL"
typ| n
---| ---
BF | 22
IC | 11
OT | 20
SP | 121
```

Butterflies (spreads of spreads)

Outrights ("normal" futures)

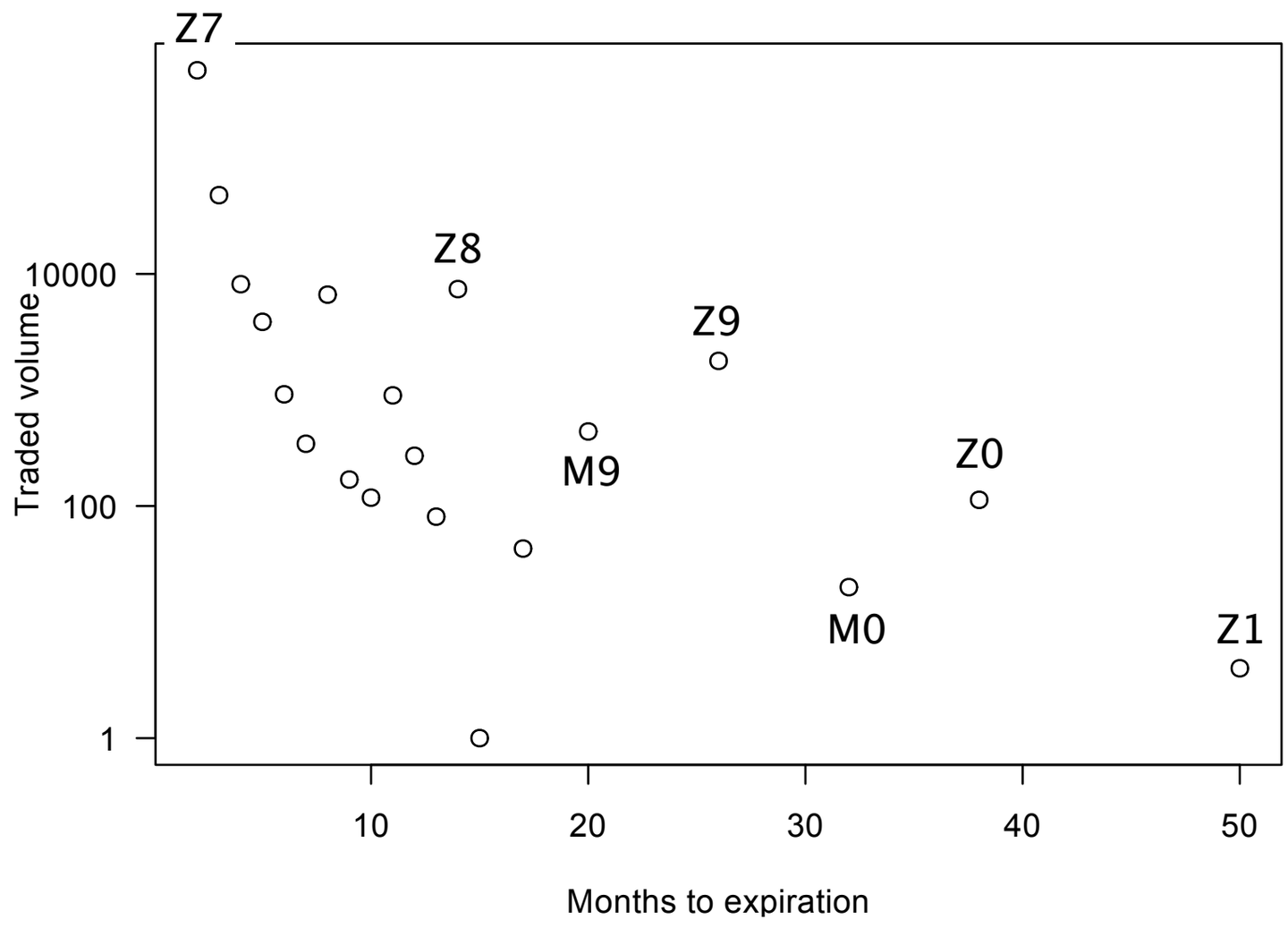
Calendar spreads

Distribution of traded volume for CL outrights

Add the expiration using `sym2expir` and sort by that

Select only outright symbols (you almost always want to do this)

```
q)h"`expir xasc update expir:sym2expir[sym] from
        select v:sum siz by sym from trade where date=2017.10.24,sym2inst[sym]=`CL,sym2type[sym]=`OT"
sym | v      expir
----| --------------
CLZ7| 568398 2017.12
CLF8| 47706  2018.01
CLG8| 8156   2018.02
CLH8| 3864   2018.03
CLJ8| 917    2018.04
CLK8| 344    2018.05
CLM8| 6632   2018.06
CLN8| 169    2018.07
CLQ8| 118    2018.08
CLU8| 898    2018.09
CLV8| 271    2018.10
CLX8| 81     2018.11
CLZ8| 7399   2018.12
CLF9| 1      2019.01
CLH9| 43     2019.03
CLM9| 439    2019.06
CLZ9| 1780   2019.12
CLM0| 20     2020.06
CLZ0| 113    2020.12
CLZ1| 4      2021.12
```



Q table syntax and joins

- Table creation (mostly for toy examples
- Keyed tables

```
q)t:([] x: 1 2 2 3; y:`a`b`c`d )

q)t
x y
---
1 a
2 b
2 c
3 d

q)select last y by x from t
x| y
-| -
1| a
2| c
3| d
```

- lj left join
- aj as-of join

`lj` "left join"

- match on shared column
- right table must have key

```
q)t
x y
---
1 a
2 b
2 c
3 d

q)t1: ([] y:`a`c; z: 3.14 2.71 )

q)select last z by y from t1
y| z
-| ----
a| 3.14
c| 2.71

q)t lj select last z by y from t1
x y z
--------
1 a 3.14
2 b
2 c 2.71
3 d
```

lj example

```
q)h1:hopen `:orf474:6005

q)h1"([] inst:insts) lj select last name by inst from instinfo"
inst name
-------------------------------------
ZN   10 Yr Note
GE   Eurodollar
ES   E-mini S&P 500
ZF   5 Yr Note
CL   Crude Oil
GC   Gold
ZT   2 Yr Note
NQ   E-mini NASDAQ 100
ZB   U.S. Treasury Bond
6E   Euro FX
NG   Natural Gas (Henry Hub) Physical
6J   Japanese Yen
ZC   Corn
6B   British Pound
ZS   Soybean
ZQ   30 Day Federal Funds
6A   Australian Dollar
TN   Ultra 10 Yr Notes
YM   E-mini Dow ($5)
UB   Ultra U.S. Treasury Bond
HG   Copper
SI   Silver
RTY  R2000 Ind Mini
6C   Canadian Dollar
ZW   Wheat
ZL   Soybean Oil
6M   Mexican Peso
ZM   Soybean Meal
RB   RBOB Gasoline Physical
HO   NY Harbor ULSD
```

`aj` "as-of join"

typically

```
aj[ `seq; (trade table); (quote table) ]
```

```
q)t:([] s: `float$1 5 6 7 9)
q)t
s
-
1
5
6
7
9
```

```
q)q:([] s:asc 5 ? 10.; b: 5 ? 20 )
q)q
s              b
---------------
2.37128796056 14
4.26917694043 17
5.67081043962 14
7.43328513578 17
9.99608161626 16
```

```
q)aj[ `s; t; q ]
s b
----
1
5 17
6 14
7 14
9 17
```

Q functions

```
q)f:{[x;y] x + 2*y }

q)f[1;2]
5
```

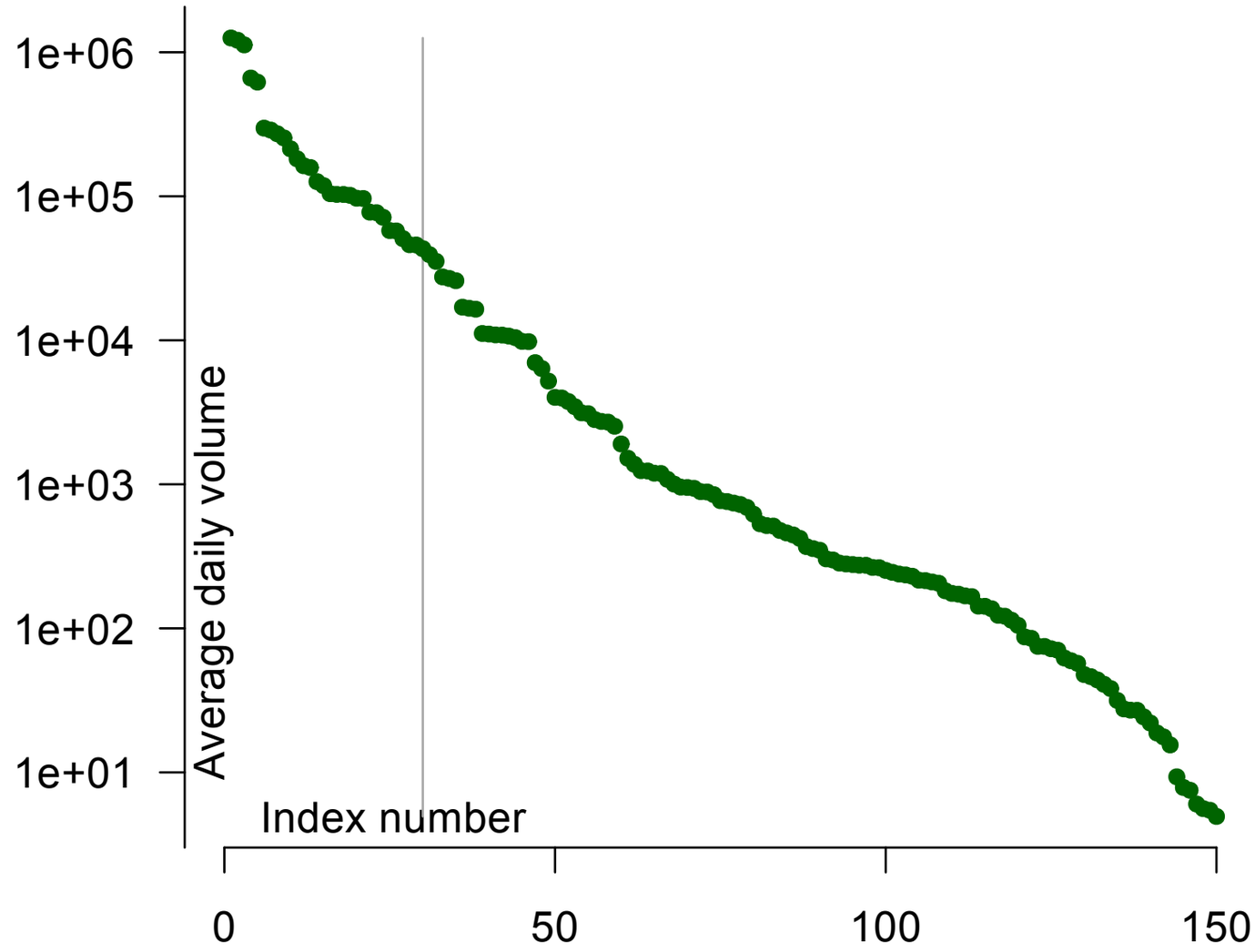
Function of 2 variables

```
q)g:f[1]

q)g 2
5

q)g 4
9
```

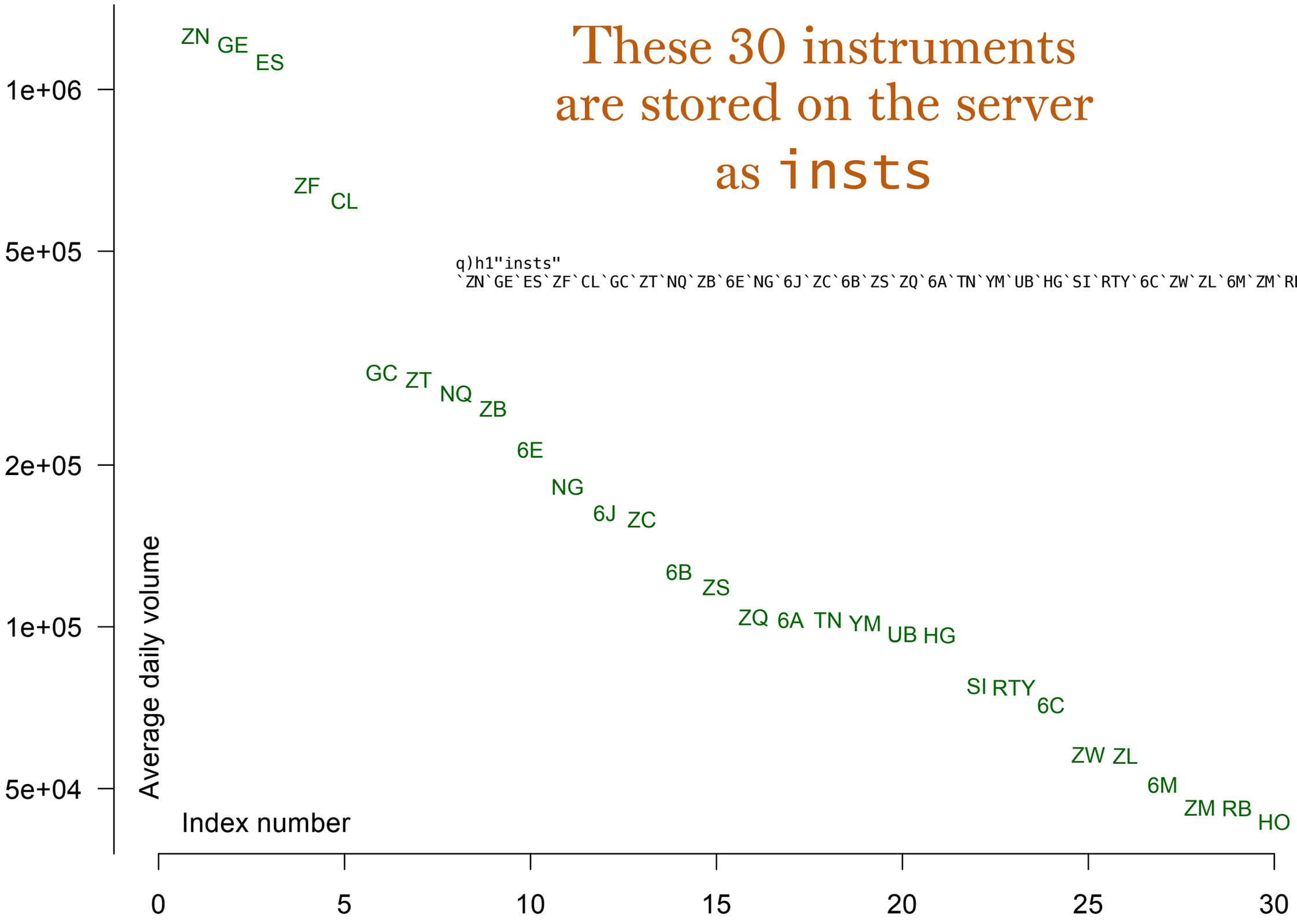
$g[y] = f[1;y]$

Most active instruments

```
q)h1"(`v xdesc select v:sum siz by inst:sym2inst[sym] from trade
        where date=2017.10.25,sym2type[sym]=`OT)
    lj select last name by inst from instinfo"
inst| v       name
----| ----------------------------------------------
ZN  | 1878740 10 Yr Note
GE  | 1480114 Eurodollar
ZF  | 962793  5 Yr Note
ZB  | 357519  U.S. Treasury Bond
ZT  | 312571  2 Yr Note
TN  | 147786  Ultra 10 Yr Notes
UB  | 125343  Ultra U.S. Treasury Bond
ZQ  | 81165   30 Day Federal Funds
...

q)h2"(`v xdesc select v:sum siz by inst:
        where date=2017.10.25,sym2type[s
    lj select last name by inst from ins
inst| v       name
----| ----------------------------------
ES  | 2084658 E-mini S&P 500
CL  | 567716  Crude Oil
NQ  | 435947  E-mini NASDAQ 100
GC  | 365748  Gold
6J  | 205358  Japanese Yen
6E  | 204538  Euro FX
ZC  | 196873  Corn
YM  | 188698  E-mini Dow ($5)
...
```



30 most active instruments



These 30 instruments are stored on the server as `insts`

```
q)h1"insts"
`ZN`GE`ES`ZF`CL`GC`ZT`NQ`ZB`6E`NG`6J`ZC`6B`ZS`ZQ`6A`TN`YM`UB`HG`SI`RTY`6C`ZW`ZL`6M`ZM`RB`HO
```

"Interesting" products

`Notional` = dollar value of unit change in contract price

`Minpxincr` = "tick"

`dispfactor` = price multiplier

```
q)h"insts"
`ZC`LE`ZS`ZW`CL`HO`NG`RB`ES`NQ`YM`6A`6B`6C`6E`6J`GE`ZF`ZT`ZN`UB`ZB`GC`HG`SI

q)h1"select from instinfo where inst in insts"
inst class name                                subclass          cur notional minpxincr minpxincrval dispfactor matchalgo
-------------------------------------------------------------------------------------------------------------------------
ZC   AG    Corn                                Grain And Oilseed USD 50       0.25      12.5         1          K
ZL   AG    Soybean Oil                         Grain And Oilseed USD 6        1         6            0.01       K
ZM   AG    Soybean Meal                        Grain And Oilseed USD 10       1         10           0.1        K
ZS   AG    Soybean                             Grain And Oilseed USD 50       0.25      12.5         1          K
ZW   AG    Wheat                               Grain And Oilseed USD 50       0.25      12.5         1          K
CL   EN    Crude Oil                           Crude Oil         USD 10       1         10           0.01       F
NG   EN    Natural Gas (Henry Hub) Physical    Natural Gas       USD 10       1         10           0.001      F
HO   EN    NY Harbor ULSD                      Refined Products  USD 4.2      1         4.2          0.0001     F
RB   EN    RBOB Gasoline Physical              Refined Products  USD 4.2      1         4.2          0.0001     F
RTY  EQ    R2000 Ind Mini                                        USD 0.5      10        5            0.01       F
ES   EQ    E-mini S&P 500                      US Index          USD 0.5      25        12.5         0.01       F
NQ   EQ    E-mini NASDAQ 100                   US Index          USD 0.2      25        5            0.01       F
YM   EQ    E-mini Dow ($5)                     US Index          USD 5        1         5            1          F
6M   FX    Mexican Peso                        Emerging Market   USD 0.5      10        5            1e-06      F
6A   FX    Australian Dollar                   Majors            USD 10       1         10           0.0001     F
6B   FX    British Pound                       Majors            USD 6.25     1         6.25         0.0001     F
6C   FX    Canadian Dollar                     Majors            USD 10       0.5       5            0.0001     F
6E   FX    Euro FX                             Majors            USD 12.5     0.5       6.25         0.0001     F
6J   FX    Japanese Yen                        Majors            USD 12.5     0.5       6.25         1e-06      F
GE   IR    Eurodollar                          Stirs             USD 25       0.5       12.5         0.01       A
ZQ   IR    30 Day Federal Funds                Stirs             USD 41.67    0.5       20.835       0.01       K
TN   IR    Ultra 10 Yr Notes                   US Treasury       USD 1000     0.015625  15.625       1          F
UB   IR    Ultra U.S. Treasury Bond            US Treasury       USD 1000     0.03125   31.25        1          F
ZB   IR    U.S. Treasury Bond                  US Treasury       USD 1000     0.03125   31.25        1          F
ZF   IR    5 Yr Note                           US Treasury       USD 1000     0.0078125 7.8125       1          F
ZN   IR    10 Yr Note                          US Treasury       USD 1000     0.015625  15.625       1          F
ZT   IR    2 Yr Note                           US Treasury       USD 2000     0.0078125 15.625       1          K
HG   MT    Copper                              Base              USD 2.5      5         12.5         0.0001     F
GC   MT    Gold                                Precious          USD 10       1         10           0.1        F
SI   MT    Silver                              Precious          USD 5        5         25           0.001      F
```

Scale factors: example

```
inst class name                                subclass          cur notional minpxincr minpxincrval dispfactor matchalgo
-----------------------------------------------------------------------------------------------------------------------
6C    FX    Canadian Dollar                    Majors            USD 10       0.5       5            0.0001     F
```

```
q)h"select v:sum siz by prc from trade where date=2017.10.26,sym=`6CZ7"
prc    | v
-------| ----
7780   | 8
7780.5| 66
7781   | 360
7781.5| 300
7782   | 485
7782.5| 665
7783   | 848
7783.5| 747
7784   | 903
7784.5| 1220
7785   | 1093
7785.5| 1293
..
```

minpxincr=0.5

Notional = USD 10:
Unit change in contract price
is worth $10

Dispfactor=0.0001: CAD/USD = 0.7782, etc

$$\Delta p = 1$$
$$= 1e\text{-}4 \text{ on } 1e5 \text{ CAD}$$
$$= \$10 \text{ USD}$$

Canadian Dollar Futures Contract Specs

View Another Product

Quotes | Settlements | Volume | Time & Sales | Contract Specs | Margins | Calendar

Futures | Options

Contract Unit	100,000 Canadian dollars
Trading Hours	Sunday - Friday 6:00 p.m. - 5:00 p.m. (5:00 p.m. - 4:00 p.m. Chicago Time/CT) with a 60-minute break each day beginning at 5:00 p.m. (4:00 p.m. CT)
Minimum Price Fluctuation	Outrights: .00005 USD per CAD ($5.00 USD). Consecutive Month Spreads (Globex only): 0.00001 USD per CAD (1.00 USD) All other Spread Combinations: 0.00005 USD per CAD (5.00 USD).
Product Code	CME Globex: 6C CME ClearPort: C1 Clearing: C1
Listed Contracts	Contracts listed for the first 3 consecutive months and 20 months in the March quarterly cycle (Mar, Jun, Sep, Dec)
Settlement Method	Deliverable
Termination Of Trading	9:16 a.m. Central Time (CT) on the business day immediately preceding the third Wednesday of the contract month (usually Tuesday).
Settlement Procedures	Physical Delivery Canadian Dollar Settlement Procedures

Match algorithms

```
q)h2"(`matchalgo xasc select inst,name,matchalgo from instinfo where inst in insts) lj matchalgos"
inst name                                matchalgo algoname
-----------------------------------------------------------------------
GE   Eurodollar                          A         Allocation
CL   Crude Oil                           F         FIFO
NG   Natural Gas (Henry Hub) Physical    F         FIFO
HO   NY Harbor ULSD                      F         FIFO
RB   RBOB Gasoline Physical              F         FIFO
RTY  R2000 Ind Mini                      F         FIFO
ES   E-mini S&P 500                      F         FIFO
NQ   E-mini NASDAQ 100                   F         FIFO
YM   E-mini Dow ($5)                     F         FIFO
6M   Mexican Peso                        F         FIFO
6A   Australian Dollar                   F         FIFO
6B   British Pound                       F         FIFO
6C   Canadian Dollar                     F         FIFO
6E   Euro FX                             F         FIFO
6J   Japanese Yen                        F         FIFO
TN   Ultra 10 Yr Notes                   F         FIFO
UB   Ultra U.S. Treasury Bond            F         FIFO
ZB   U.S. Treasury Bond                  F         FIFO
ZF   5 Yr Note                           F         FIFO
ZN   10 Yr Note                          F         FIFO
HG   Copper                              F         FIFO
GC   Gold                                F         FIFO
SI   Silver                              F         FIFO
ZC   Corn                                K         Split FIFO and Pro-Rata
ZL   Soybean Oil                         K         Split FIFO and Pro-Rata
ZM   Soybean Meal                       K         Split FIFO and Pro-Rata
ZS   Soybean                            K         Split FIFO and Pro-Rata
ZW   Wheat                              K         Split FIFO and Pro-Rata
ZQ   30 Day Federal Funds                K         Split FIFO and Pro-Rata
ZT   2 Yr Note                          K         Split FIFO and Pro-Rata
```

Pro rata

Time priority

Mixed

Most active sym for each "interesting" contract

```
q)h1:hopen `:orf474:6003
q)h2:hopen `:orf474:7003

q)h1"select first sym,first v by inst from
      update inst:sym2inst[sym] from
       `v xdesc
         select v:sum siz by sym from trade where date=2017.10.25,sym2inst[sym] in insts,sym2type[sym]=`OT"
inst| sym  v
----| ------------
GE  | GEZ8 206841
TN  | TNZ7 147786
UB  | UBZ7 125342
ZB  | ZBZ7 357455
ZF  | ZFZ7 962785
ZN  | ZNZ7 1878572
ZQ  | ZQF8 18482
ZT  | ZTZ7 311933

q)h2"select first sym,first v by inst from
      update inst:sym2inst[sym] from
       `v xdesc
        select v:sum siz by sym from trade where date=2017.10.25,sym2inst[sym] in insts,sym2type[sym]=`OT"
inst| sym   v
----| -------------
6A  | 6AZ7  145711
6B  | 6BZ7  151376
6C  | 6CZ7  136228
6E  | 6EZ7  202545
6J  | 6JZ7  204642
6M  | 6MZ7  80084
CL  | CLZ7  506586
ES  | ESZ7  2082705
GC  | GCZ7  363223
HG  | HGZ7  100796
HO  | HOZ7  35471
NG  | NGZ7  82811
NQ  | NQZ7  435242
RB  | RBZ7  37096
RTY | RTYZ7 95737
SI  | SIZ7  88525
YM  | YMZ7  188541
ZC  | ZCZ7  141346
ZL  | ZLZ7  44246
ZM  | ZMZ7  35141
ZS  | ZSX7  54143
ZW  | ZWZ7  42794
```

For almost all products,
the first quarterly expiration
is the most active
(Dec 2017 = Z7, in Oct 2017)

When we talk about "the" symbol
for a particular instrument,
we mean the symbol that is
most actively traded

"Entropy" and "Effective number"

```
q)h2"`v xdesc select v:sum siz by sym from trade where date=2017.10.25,sym2inst[sym]=`ES,sym2type[sym]=`OT"
sym | v
----| -------
ESZ7| 2082705
ESH8| 1903
ESM8| 37
ESU8| 13

q)h2"`v xdesc select v:sum siz by sym from trade where date=2017.10.25,sym2inst[sym]=`CL,sym2type[sym]=`OT"
sym | v
----| ------
CLZ7| 506586
CLF8| 42942
CLG8| 4990
CLZ8| 4021
CLM8| 3065
CLH8| 2737
CLJ8| 1000
CLZ9| 956
CLK8| 400
CLN8| 308
CLM9| 277
CLU8| 199
CLQ8| 101
CLZ0| 91
CLV8| 23
CLX8| 13
CLZ1| 5
CLM0| 1
CLZ2| 1
```

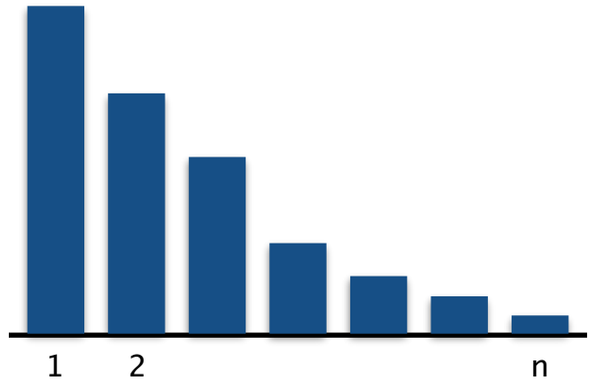
ES has "few" active symbols

CL has "many" active symbols
but how many?

Should we count these tiny ones?

Diversity measures

Given $p_1, \ldots, p_n$ with $p_1 + \cdots + p_n = 1$



What is the "effective number" of values?

Ecology -- number of species in an environment

- Simpson index

Antitrust -- number of companies in an industry

- Herfindahl-Hirschman index

Signal processing -- information content

- Shannon entropy

Futures markets -- characterization of products

Antitrust

THE UNITED STATES DEPARTMENT of JUSTICE

HERFINDAHL-HIRSCHMAN INDEX

The term “HHI” means the Herfindahl–Hirschman Index, a commonly accepted measure of market concentration. The HHI is calculated by squaring the market share of each firm competing in the market and then summing the resulting numbers. For example, for a market consisting of four firms with shares of 30, 30, 20, and 20 percent, the HHI is 2,600 ($30^2 + 30^2 + 20^2 + 20^2 = 2{,}600$).

The HHI takes into account the relative size distribution of the firms in a market. It approaches zero when a market is occupied by a large number of firms of relatively equal size and reaches its maximum of 10,000 points when a market is controlled by a single firm. The HHI increases both as the number of firms in the market decreases and as the disparity in size between those firms increases.

DIVERSITY AND EVENNESS: A UNIFYING NOTATION AND ITS CONSEQUENCES[1]

M. O. HILL[2]

School of Plant Biology, University College of North Wales, Bangor, Caernarvonshire, Wales

Abstract. Three commonly used measures of diversity, Simpson's index, Shannon's entropy, and the total number of species, are related to Rényi's definition of a generalized entropy. A unified concept of diversity is presented, according to which there is a continuum of possible diversity measures. In a sense which becomes apparent, these measures provide estimates of the effective number of species present, and differ only in their tendency to include or to ignore the relatively rarer species. The notion of the diversity of a community as opposed to that of a sample is examined, and is related to the asymptotic form of the species-abundance curve. A new and plausible definition of evenness is derived.

When we say that the humid tropics are more diverse than the tundra, we mean that there are more species there. More precisely, we mean that the species in the humid tropics have on average lower proportional abundances than those in the tundra—a fact which is amply visible to the naked eye and which can be demonstrated by the use of any measure of diversity we care to devise. But there is little point in merely confirming the obvious; the purpose of determining diversity by a numerical index is rather to provide a means of comparison between less clear-cut cases. Unfortunately, when we look for a suitable numerical definition, we find that no particular formula has a pre-eminent advantage, and that different authors have plausibly proposed different indices. In the ensuing confusion, Hurlbert (1971) has despaired, declaring diversity to be a nonconcept. Fortunately his despair is premature, and when carefully defined according to an appropriate notation, diversity can be as unequivocal as any other ecological parameter.

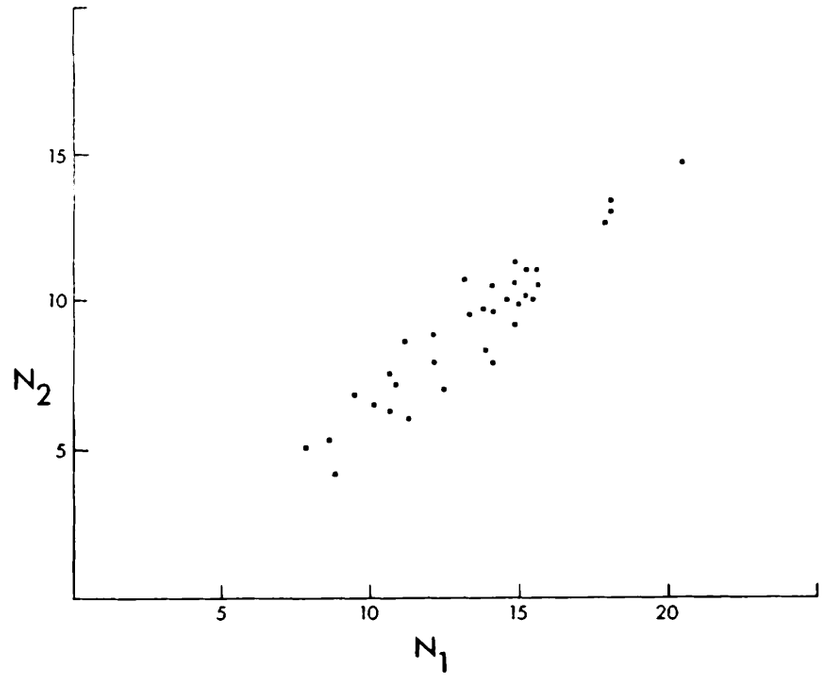


FIG. 3. Scatter diagram of N_1 versus N_2, computed for 30 cm × 30 cm dry-weight samples in a pasture in North Wales.

A Mathematical Theory of Communication

By C. E. SHANNON

INTRODUCTION

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist[1] and Hartley[2] on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.

6. CHOICE, UNCERTAINTY AND ENTROPY

We have represented a discrete information source as a Markoff process. Can we define a quantity which will measure, in some sense, how much information is "produced" by such a process, or better, at what rate information is produced?

Suppose we have a set of possible events whose probabilities of occurrence are $p_1, p_2, \ldots, p_n$. These probabilities are known but that is all we know concerning which event will occur. Can we find a measure of how much "choice" is involved in the selection of the event or of how uncertain we are of the outcome?

Theorem 2: The only H satisfying the three above assumptions is of the form:

$$H = -K \sum_{i=1}^{n} p_i \log p_i$$

where K is a positive constant.

Reprinted with corrections from *The Bell System Technical Journal,*
Vol. 27, pp. 379–423, 623–656, July, October, 1948.

Mean of $(x_1, \ldots, x_n)$ under $(p_1, \ldots, p_n)$

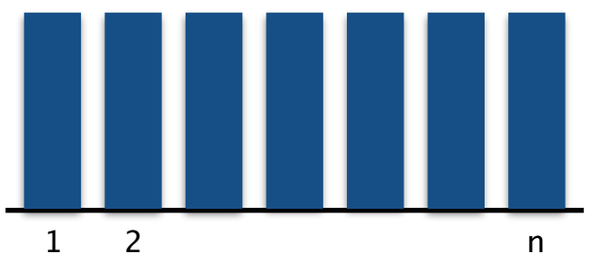
$$\langle x \rangle \;=\; \sum p_j \, x_j$$

Generalized mean of $(x_1, \ldots, x_n)$ under $(p_1, \ldots, p_n)$

$$M_q(x) \;=\; \langle x^q \rangle^{1/q} \;=\; \left(\sum p_j \, x_j^q \right)^{1/q}$$

Effective number of order q

$$N_q(p) \;=\; \frac{1}{M_q(p)} \;=\; \frac{1}{\left(\sum p_j^{q+1} \right)^{1/q}}$$



$$p_j = \frac{1}{n} \quad \Longrightarrow \quad N_q(p) = n$$

Shannon entropy

$$N_0(p) \;=\; \exp \sum -p_j \log p_j$$

Simpson index (ecology)
Herfindahl-Hirschman index (economics)

$$N_1(p) \;=\; \frac{1}{\sum p_j^2}$$

$$N_\infty(p) \;=\; \frac{1}{\max p_j}$$

Apply to a single instrument

```
q)h2"{[dmin;dmax;inst]
    exec exp sum neg p*log p from
    update p:v % sum v from
    select v:sum siz by sym from trade where date within (dmin;dmax),
             sym2inst[sym]=inst,sym2type[sym]=`OT
       }[2017.10.01;2017.10.31;`CL]"

2.8131619113
```

Apply to multiple instruments

```
q)h2"{[dmin;dmax;inst]
     exec exp sum neg p*log p from
     update p:v % sum v from
     select v:sum siz by sym from trade where date within (dmin;dmax),
              sym2inst[sym]=inst,sym2type[sym]=`OT
          }[2017.10.01;2017.10.31] each `CL`ES"

2.8131619113 1.00890548575
```

Apply to all instruments

```
q)h2"`Neff xasc ([] inst:insts;
    Neff:{[dmin;dmax;inst]
       exec exp sum neg p*log p from update p:v % sum v from
        select v:sum siz by sym from trade where date within (dmin;dmax),
           sym2inst[sym]=inst,sym2type[sym]=`OT}[2017.10.01;2017.10.31]
    each insts)"

inst Neff
------------------
RTY  1.00034196781
6M   1.00452172163
ES   1.00890548575
YM   1.0097401893
NQ   1.01294970827
6J   1.02330012816
6B   1.02431960184
6A   1.03225770414
6C   1.03362941301
GC   1.05753871273
SI   1.06632280704
6E   1.07518727869
HG   1.2257078794
ZW   1.98039363339
ZC   2.31684118338
ZM   2.46968492734
ZL   2.47885402264
CL   2.8131619113
ZS   2.875605368
RB   3.51759025319
NG   3.6786605864
HO   3.68320053807
```

Most products have one active maturity

```
q)h1"`Neff xasc ([] inst:insts;
    Neff:{[dmin;dmax;inst]
       exec exp sum neg p*log p from update p:v % sum v from
        select v:sum siz by sym from trade where date within (dmin;dmax),
           sym2inst[sym]=inst,sym2type[sym]=`OT}[2017.10.01;2017.10.31]
    each insts)"
inst Neff
------------------
UB   1.00009914399
ZF   1.00019282326
ZN   1.00136113178
ZB   1.00166131796
ZT   1.0019205291
ZQ   11.422332455
GE   15.6192927728
```

Most products have one active maturity

A few products (energy, agriculturals, Eurodollars) have multiple active maturities

Entropy and Futures Contracts

Isaac Carruthers

June 29, 2015