JD Bitcoin Blockchain Spelunking

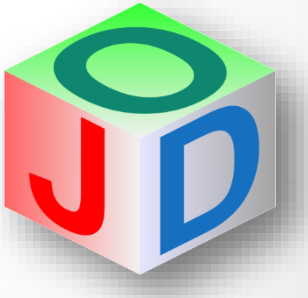
Who am I?

Does it really matter?

John D. Baker

- Yet another corporate programmer – working in insurance
- Longtime J user -- see JOD addon & Github jacks repository
- Lackadaisical Blogger -- see Analyze the Data not the Drivel
- Skeptical curmudgeon with overt libertarian tendencies

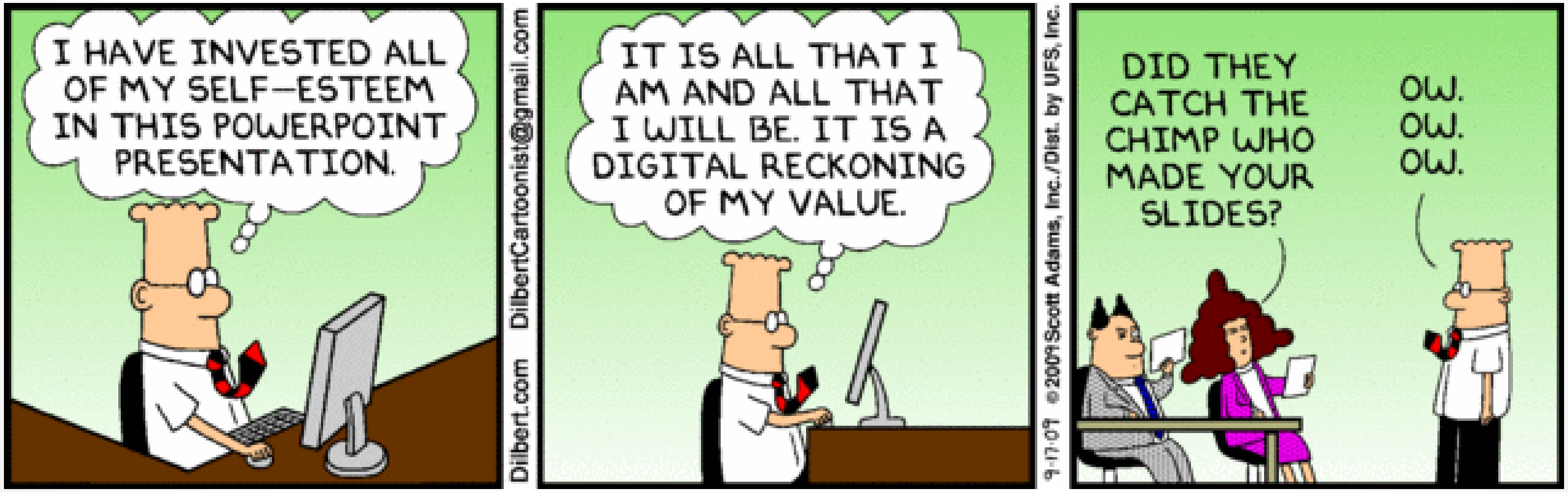
and finally:



Bitcoin curious!

Motivations

- Experiment with JD
- Learn more about Bitcoin
- Avoid death by PowerPoint


I HAVE INVESTED ALL OF MY SELF—ESTEEM IN THIS POWERPOINT PRESENTATION.
IT IS ALL THAT I AM AND ALL THAT I WILL BE. IT IS A DIGITAL RECKONING OF MY VALUE.
DID THEY CATCH THE CHIMP WHO MADE YOUR SLIDES?
OW. OW. OW.
Dilbert.com DilbertCartoonist@gmail.com
9-17-09 ©2009 Scott Adams, Inc./Dist. by UFS, Inc.

What's JD?

- JD is Jsoftware's column oriented in memory (inverted) database system.
- JD is unabashedly embedded in the J runtime environment.
- JD is designed for analytics.
- JD has efficient CSV bulk load and unload.
- JD requires 64 bit versions of J and makes extensive use of file mapping.
- JD database size is primarily limited by memory.

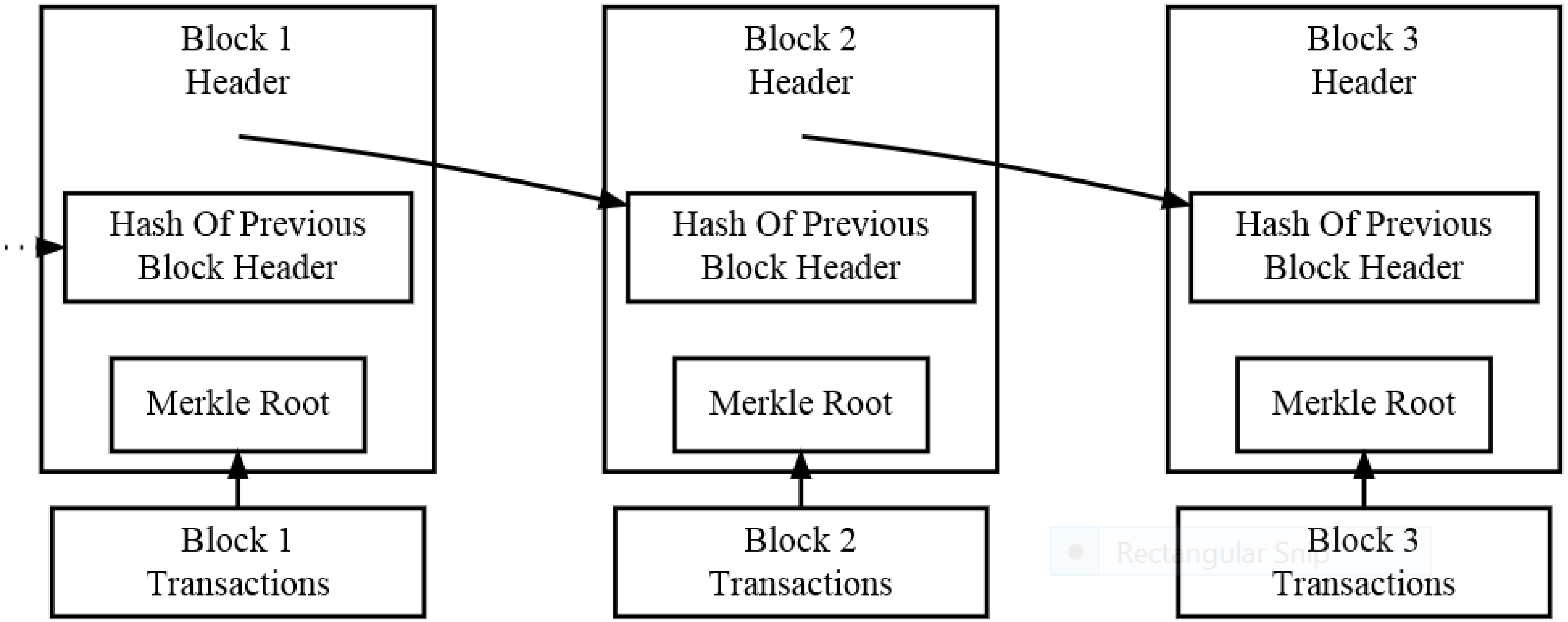
What's Bitcoin?

https://www.youtube.com/watch?v=Gc2en3nHxA4

What's a Blockchain?

- A public global cryptographically secured ledger.
- Every current full Bitcoin node and "miner" has a complete copy of the blockchain.
- The blockchain is immutable. After a block is added to the blockchain it can never be altered.
- The Bitcoin blockchain is built by competing "miners" at a rate of about one block every ten minutes.
- Each block contains one or more transactions.
- Each transaction contains one or more inputs and one or more outputs.

What's a Blockchain?

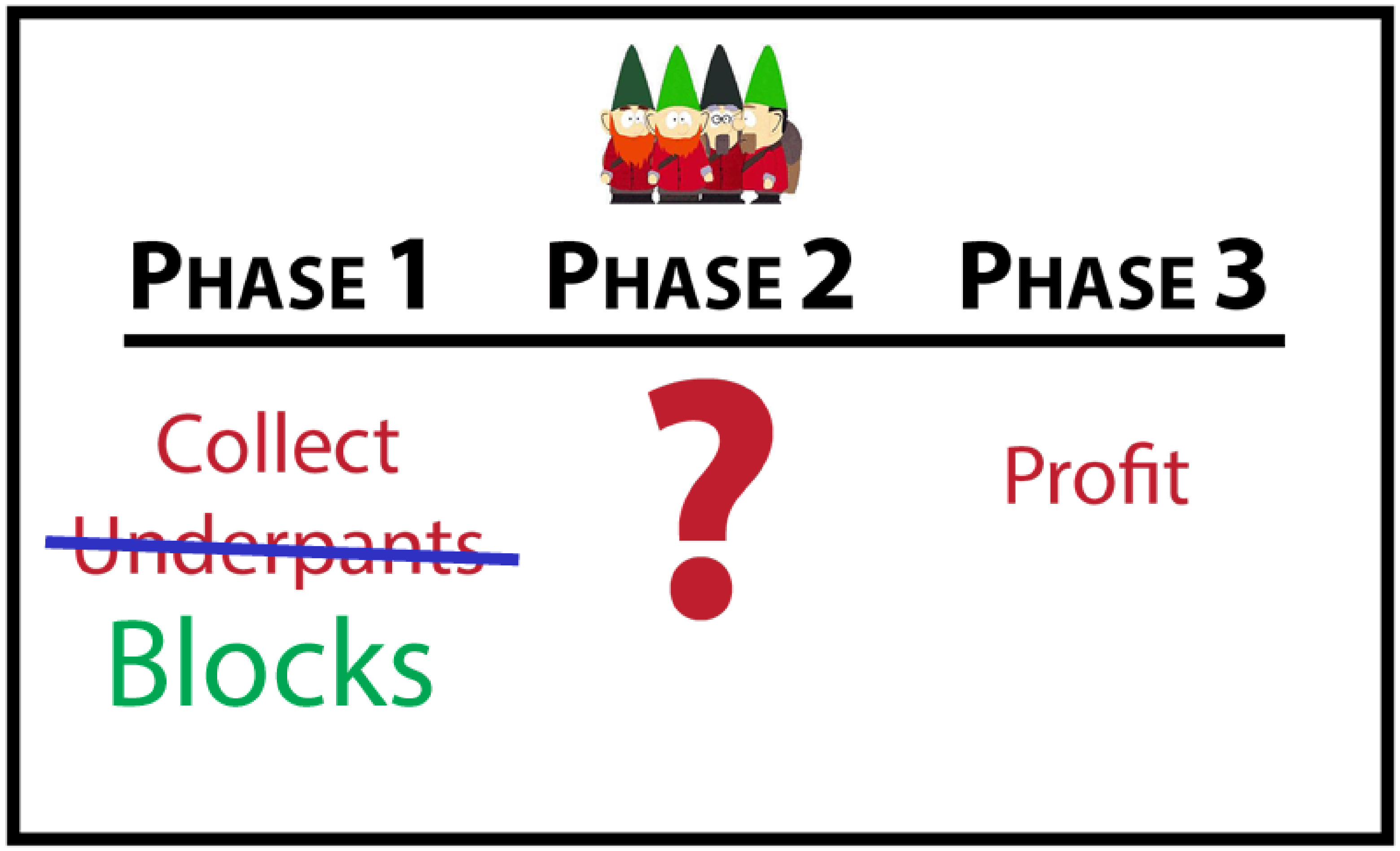


Simplified Bitcoin Block Chain

The Genesis Block

- Is the first block on the blockchain.
- It was mined on January 3, 2009 by the still mysterious Satoshi Nakamoto.
- *"The Times 03/Jan/2009 Chancellor on brink of second bailout for banks."*
- *For more see:*
  - https://en.bitcoin.it/wiki/Genesis_block
  - http://codesuppository.blogspot.com/2014/01/how-to-parse-bitcoin-blockchain.html
  - http://bakerjd99.wordpress.com/2014/07/03/parsing-the-bitcoin-genesis-block-with-j/

Blockchain into JD



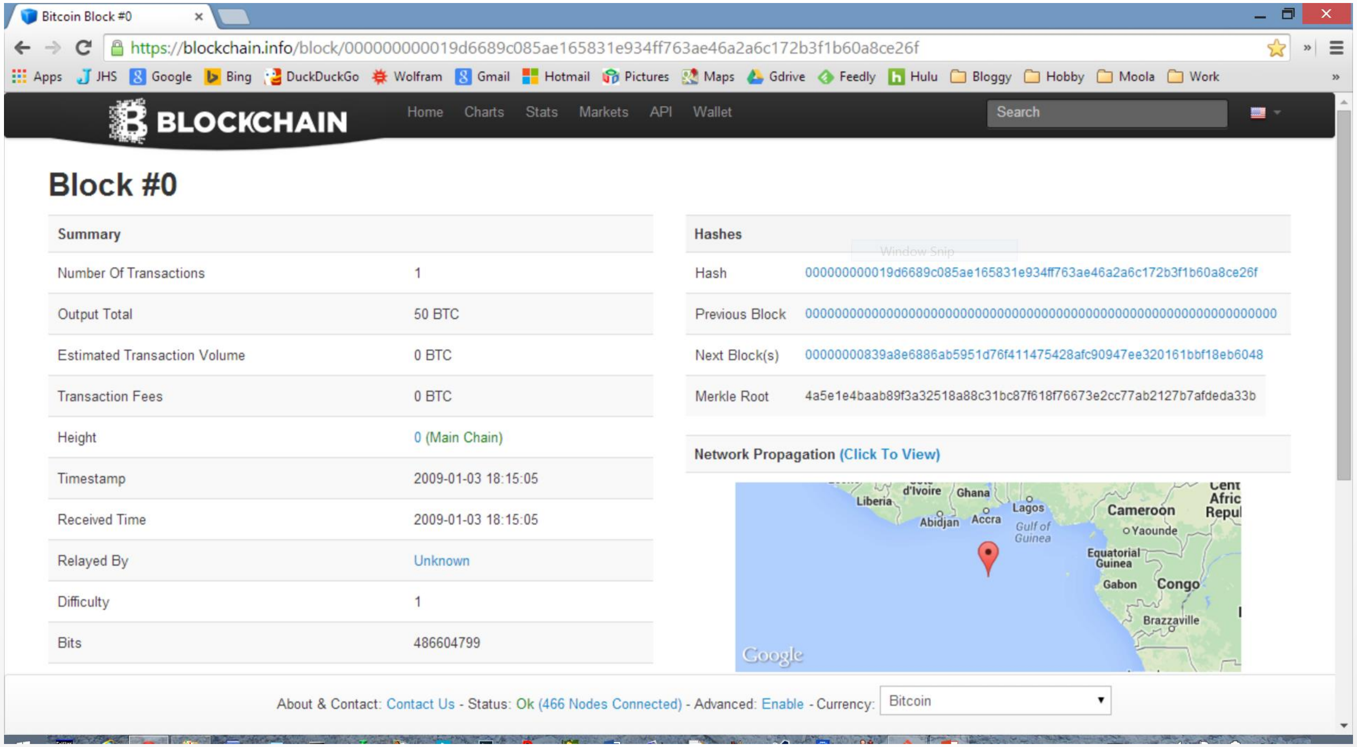
Phase 1	Phase 2	Phase 3
Collect ~~Underpants~~ Blocks	?	Profit

Blockchain into JD

PHASE 1	PHASE 2	PHASE 3
Blocks	ETL	JD/csv

Blocks

- All blockchain blocks are available online. Anyone can inspect a block. This is how many confirm transactions.



Block #0

Summary	
Number Of Transactions	1
Output Total	50 BTC
Estimated Transaction Volume	0 BTC
Transaction Fees	0 BTC
Height	0 (Main Chain)
Timestamp	2009-01-03 18:15:05
Received Time	2009-01-03 18:15:05
Relayed By	Unknown
Difficulty	1
Bits	486604799

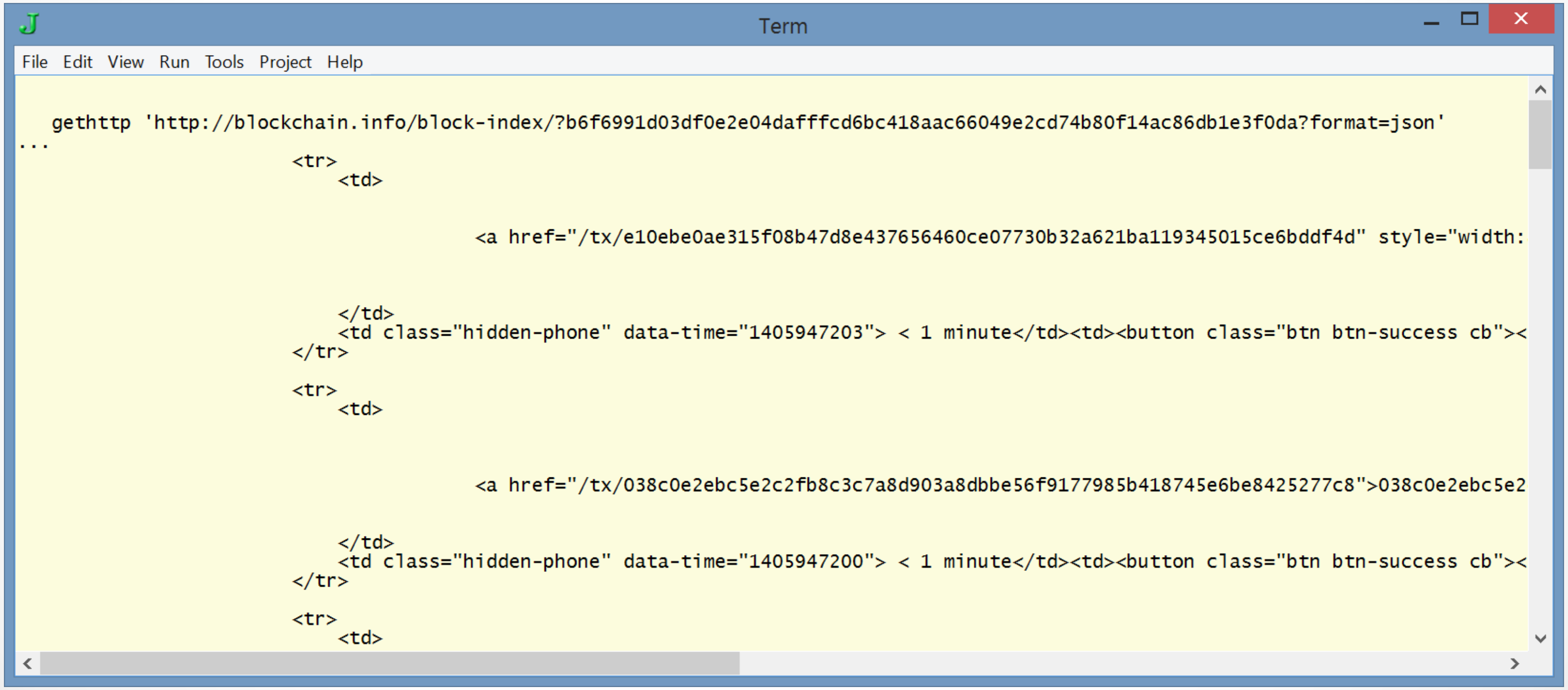
Hashes	
Hash	000000000019d6689c085ae165831e934ff763ae46a2a6c172b3f1b60a8ce26f
Previous Block	0000000000000000000000000000000000000000000000000000000000000000
Next Block(s)	00000000839a8e6886ab5951d76f411475428afc90947ee320161bbf18eb6048
Merkle Root	4a5e1e4baab89f3a32518a88c31bc87f618f76673e2cc77ab2127b7afdeda33b

Network Propagation (Click To View)

About & Contact: Contact Us - Status: Ok (466 Nodes Connected) - Advanced: Enable - Currency: Bitcoin

Blocks

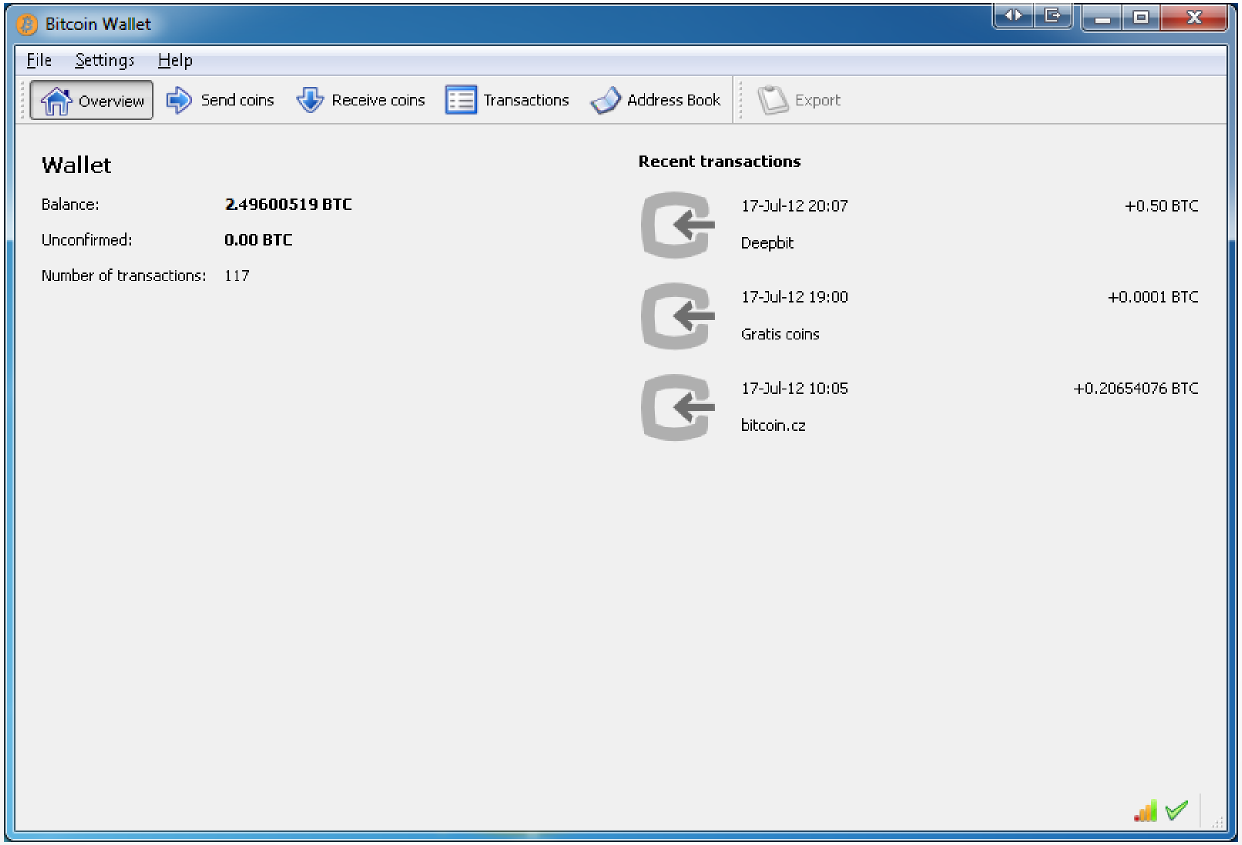
- Blocks can be fetched in many formats. JSON is popular.



```
   gethttp 'http://blockchain.info/block-index/?b6f6991d03df0e2e04dafffcd6bc418aac66049e2cd74b80f14ac86db1e3f0da?format=json'
...
                        <tr>
                            <td>


                                        <a href="/tx/e10ebe0ae315f08b47d8e437656460ce07730b32a621ba119345015ce6bddf4d" style="width:


                            </td>
                            <td class="hidden-phone" data-time="1405947203"> < 1 minute</td><td><button class="btn btn-success cb"><
                        </tr>

                        <tr>
                            <td>



                                        <a href="/tx/038c0e2ebc5e2c2fb8c3c7a8d903a8dbbe56f9177985b418745e6be8425277c8">038c0e2ebc5e2


                            </td>
                            <td class="hidden-phone" data-time="1405947200"> < 1 minute</td><td><button class="btn btn-success cb"><
                        </tr>

                        <tr>
                            <td>
```

Blocks

- Block data is maintained by the standard Bitcoin client.



Wallet

Balance: 2.49600519 BTC

Unconfirmed: 0.00 BTC

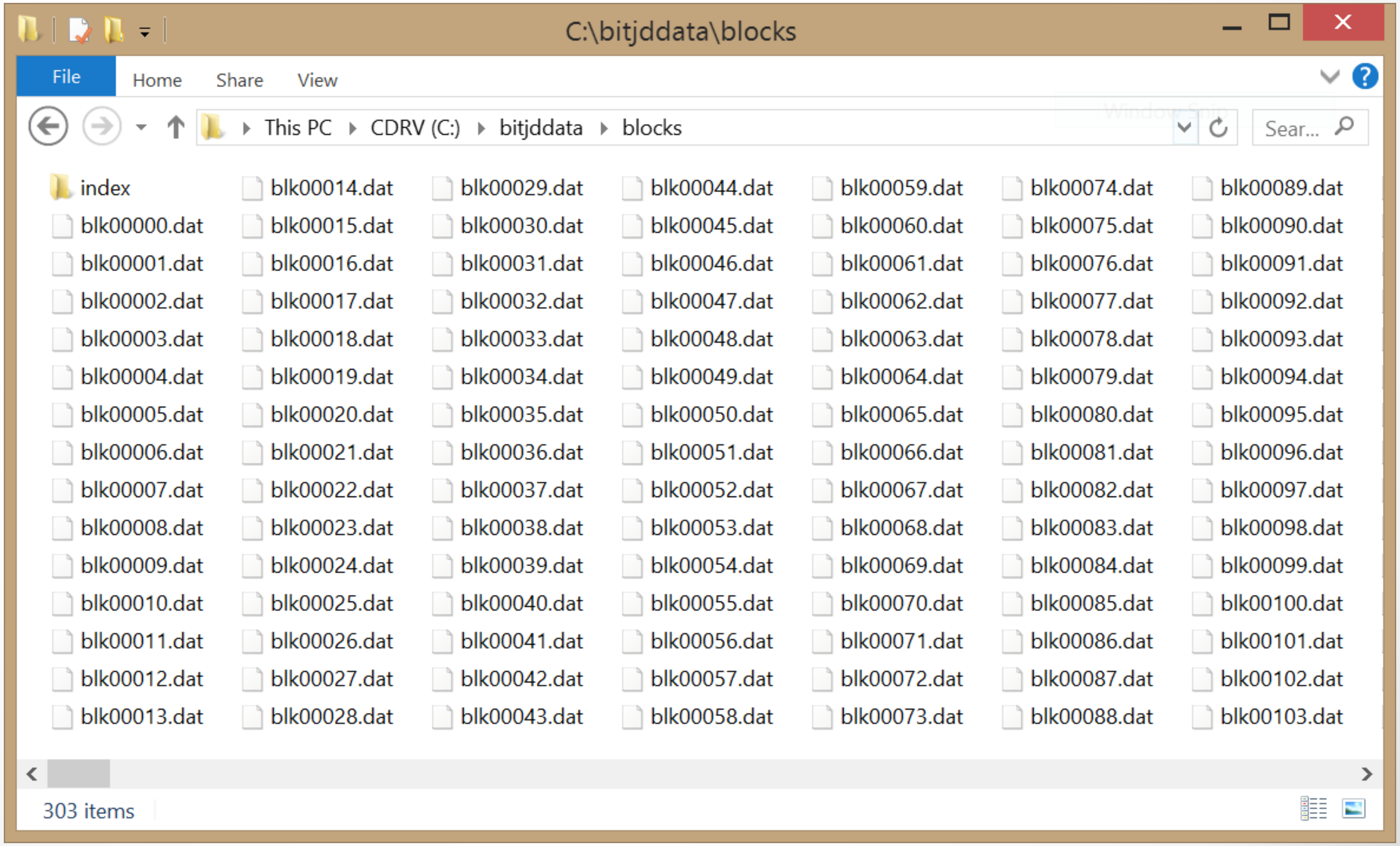
Number of transactions: 117

Recent transactions

17-Jul-12 20:07	Deepbit	+0.50 BTC
17-Jul-12 19:00	Gratis coins	+0.0001 BTC
17-Jul-12 10:05	bitcoin.cz	+0.20654076 BTC

Blocks

- Bitcoin block data is stored in a directory of binary files.



C:\bitjddata\blocks

index
blk00000.dat
blk00001.dat
blk00002.dat
blk00003.dat
blk00004.dat
blk00005.dat
blk00006.dat
blk00007.dat
blk00008.dat
blk00009.dat
blk00010.dat
blk00011.dat
blk00012.dat
blk00013.dat
blk00014.dat
blk00015.dat
blk00016.dat
blk00017.dat
blk00018.dat
blk00019.dat
blk00020.dat
blk00021.dat
blk00022.dat
blk00023.dat
blk00024.dat
blk00025.dat
blk00026.dat
blk00027.dat
blk00028.dat
blk00029.dat
blk00030.dat
blk00031.dat
blk00032.dat
blk00033.dat
blk00034.dat
blk00035.dat
blk00036.dat
blk00037.dat
blk00038.dat
blk00039.dat
blk00040.dat
blk00041.dat
blk00042.dat
blk00043.dat
blk00044.dat
blk00045.dat
blk00046.dat
blk00047.dat
blk00048.dat
blk00049.dat
blk00050.dat
blk00051.dat
blk00052.dat
blk00053.dat
blk00054.dat
blk00055.dat
blk00056.dat
blk00057.dat
blk00058.dat
blk00059.dat
blk00060.dat
blk00061.dat
blk00062.dat
blk00063.dat
blk00064.dat
blk00065.dat
blk00066.dat
blk00067.dat
blk00068.dat
blk00069.dat
blk00070.dat
blk00071.dat
blk00072.dat
blk00073.dat
blk00074.dat
blk00075.dat
blk00076.dat
blk00077.dat
blk00078.dat
blk00079.dat
blk00080.dat
blk00081.dat
blk00082.dat
blk00083.dat
blk00084.dat
blk00085.dat
blk00086.dat
blk00087.dat
blk00088.dat
blk00089.dat
blk00090.dat
blk00091.dat
blk00092.dat
blk00093.dat
blk00094.dat
blk00095.dat
blk00096.dat
blk00097.dat
blk00098.dat
blk00099.dat
blk00100.dat
blk00101.dat
blk00102.dat
blk00103.dat

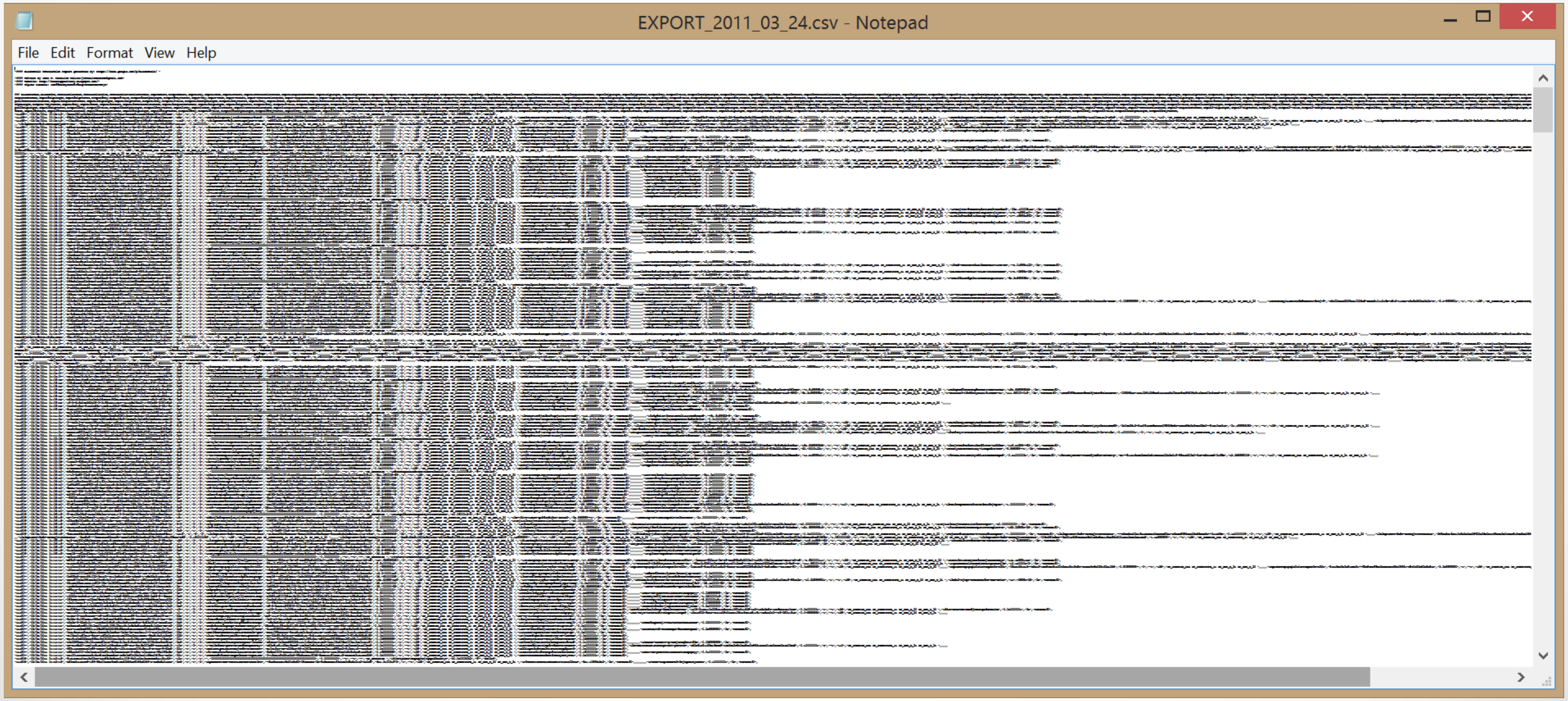
303 items

Blocks

- Block files are 10 to150 MB.
- Currently there are about 350 block files ≈ 25 GB.
- Block files cannot be directly loaded into JD.
- Convert blockchain data to CSV.
- I used John Ratcliffe's `blochain64.exe` tool to dump a sample of blockchain transactions, see:
    - http://codesuppository.blogspot.com/2014/03/how-to-extract-every-single-bitcoin.html
- The resulting CSV "transaction day" files are not normalized or ready for JD loading.

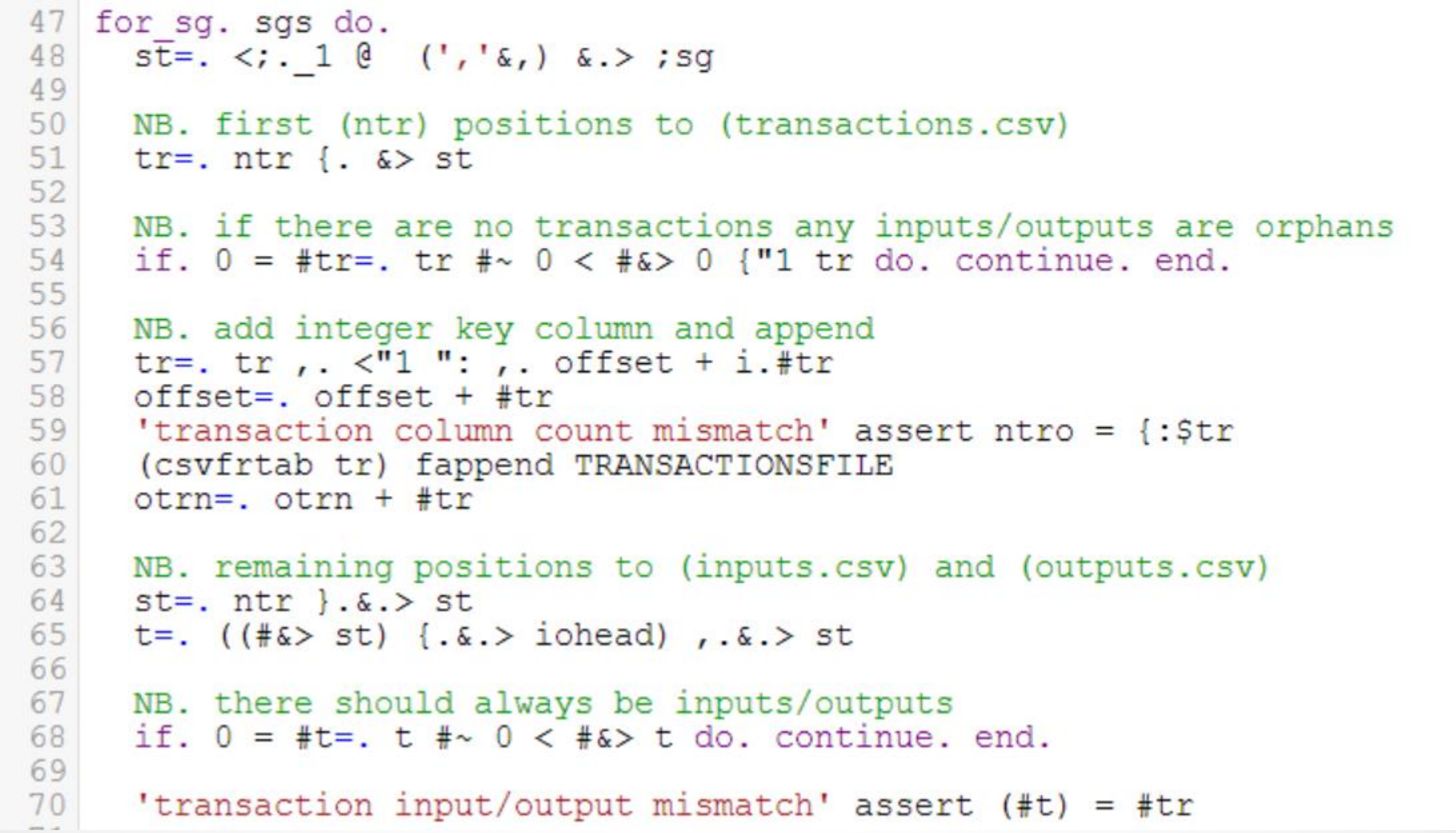
ETL

Each line in a CSV transaction day file is a single transaction with a variable number of inputs and output.



ETL

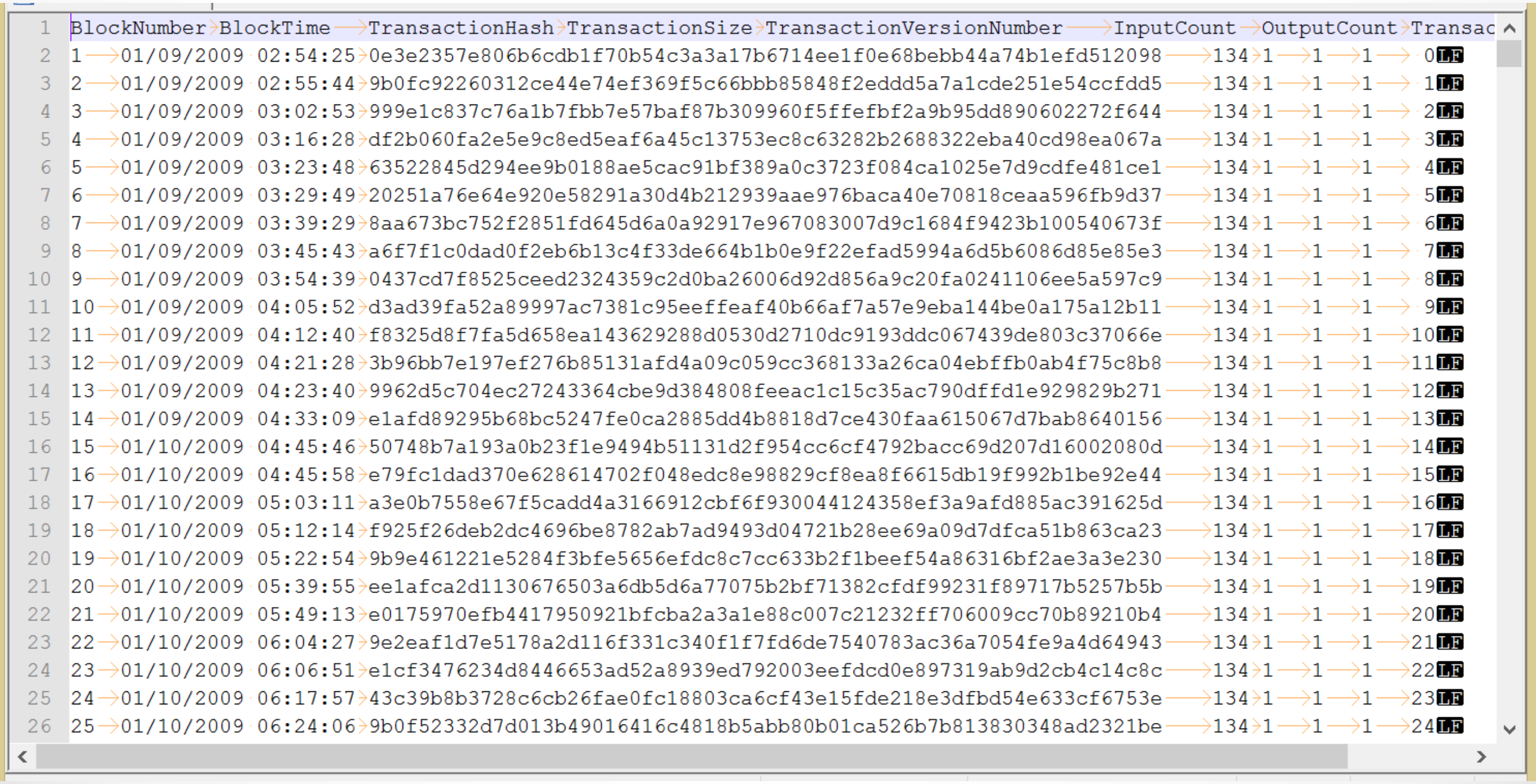
CSV files were normalized by a J script that split and appended each file to three TAB delimited CSV files.



```
for_sg. sgs do.
  st=. <;._1 @  (','&,) &.> ;sg

  NB. first (ntr) positions to (transactions.csv)
  tr=. ntr {. &> st

  NB. if there are no transactions any inputs/outputs are orphans
  if. 0 = #tr=. tr #~ 0 < #&> 0 {"1 tr do. continue. end.

  NB. add integer key column and append
  tr=. tr ,. <"1 ": ,. offset + i.#tr
  offset=. offset + #tr
  'transaction column count mismatch' assert ntro = {:$tr
  (csvfrtab tr) fappend TRANSACTIONSFILE
  otrn=. otrn + #tr

  NB. remaining positions to (inputs.csv) and (outputs.csv)
  st=. ntr }.&.> st
  t=. ((#&> st) {.&.> iohead) ,.&.> st

  NB. there should always be inputs/outputs
  if. 0 = #t=. t #~ 0 < #&> t do. continue. end.

  'transaction input/output mismatch' assert (#t) = #tr
```

ETL

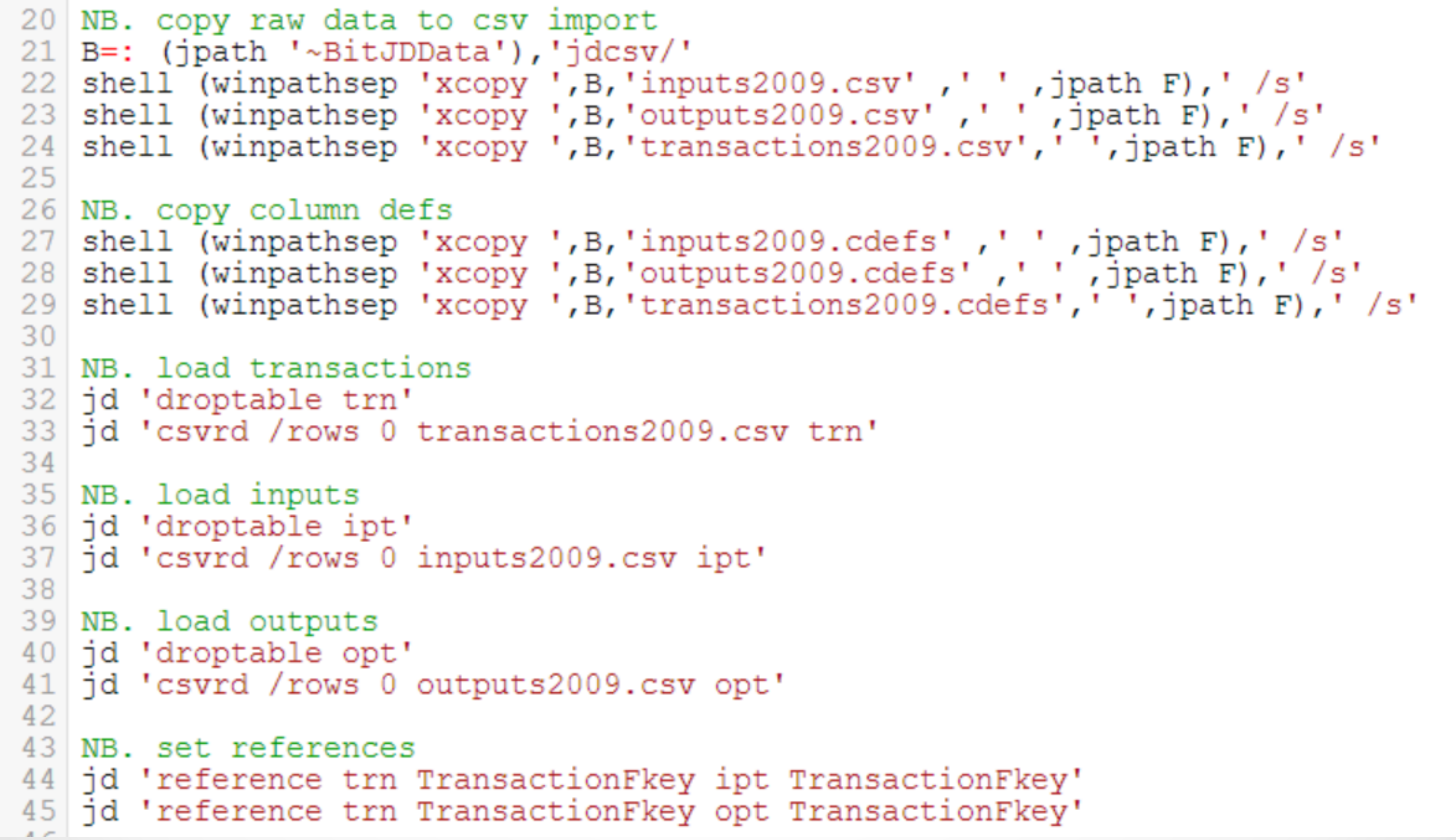
Splitting and merging the day CSV files results in regular tidy load files.



```
BlockNumber	BlockTime	TransactionHash	TransactionSize	TransactionVersionNumber	InputCount	OutputCount	Transac
1	01/09/2009 02:54:25	0e3e2357e806b6cdb1f70b54c3a3a17b6714ee1f0e68bebb44a74b1efd512098	134	1	1	1	 0
2	01/09/2009 02:55:44	9b0fc92260312ce44e74ef369f5c66bbb85848f2eddd5a7a1cde251e54ccfdd5	134	1	1	1	 1
3	01/09/2009 03:02:53	999e1c837c76a1b7fbb7e57baf87b309960f5ffefbf2a9b95dd890602272f644	134	1	1	1	 2
4	01/09/2009 03:16:28	df2b060fa2e5e9c8ed5eaf6a45c13753ec8c63282b2688322eba40cd98ea067a	134	1	1	1	 3
5	01/09/2009 03:23:48	63522845d294ee9b0188ae5cac91bf389a0c3723f084ca1025e7d9cdfe481ce1	134	1	1	1	 4
6	01/09/2009 03:29:49	20251a76e64e920e58291a30d4b212939aae976baca40e70818ceaa596fb9d37	134	1	1	1	 5
7	01/09/2009 03:39:29	8aa673bc752f2851fd645d6a0a92917e967083007d9c1684f9423b100540673f	134	1	1	1	 6
8	01/09/2009 03:45:43	a6f7f1c0dad0f2eb6b13c4f33de664b1b0e9f22efad5994a6d5b6086d85e85e3	134	1	1	1	 7
9	01/09/2009 03:54:39	0437cd7f8525ceed2324359c2d0ba26006d92d856a9c20fa0241106ee5a597c9	134	1	1	1	 8
10	01/09/2009 04:05:52	d3ad39fa52a89997ac7381c95eeffeaf40b66af7a57e9eba144be0a175a12b11	134	1	1	1	 9
11	01/09/2009 04:12:40	f8325d8f7fa5d658ea143629288d0530d2710dc9193ddc067439de803c37066e	134	1	1	1	10
12	01/09/2009 04:21:28	3b96bb7e197ef276b85131afd4a09c059cc368133a26ca04ebffb0ab4f75c8b8	134	1	1	1	11
13	01/09/2009 04:23:40	9962d5c704ec27243364cbe9d384808feeac1c15c35ac790dffd1e929829b271	134	1	1	1	12
14	01/09/2009 04:33:09	e1afd89295b68bc5247fe0ca2885dd4b8818d7ce430faa615067d7bab8640156	134	1	1	1	13
15	01/10/2009 04:45:46	50748b7a193a0b23f1e9494b51131d2f954cc6cf4792bacc69d207d16002080d	134	1	1	1	14
16	01/10/2009 04:45:58	e79fc1dad370e628614702f048edc8e98829cf8ea8f6615db19f992b1be92e44	134	1	1	1	15
17	01/10/2009 05:03:11	a3e0b7558e67f5cadd4a3166912cbf6f930044124358ef3a9afd885ac391625d	134	1	1	1	16
18	01/10/2009 05:12:14	f925f26deb2dc4696be8782ab7ad9493d04721b28ee69a09d7dfca51b863ca23	134	1	1	1	17
19	01/10/2009 05:22:54	9b9e461221e5284f3bfe5656efdc8c7cc633b2f1beef54a86316bf2ae3a3e230	134	1	1	1	18
20	01/10/2009 05:39:55	ee1afca2d1130676503a6db5d6a77075b2bf71382cfdf99231f89717b5257b5b	134	1	1	1	19
21	01/10/2009 05:49:13	e0175970efb4417950921bfcba2a3a1e88c007c21232ff706009cc70b89210b4	134	1	1	1	20
22	01/10/2009 06:04:27	9e2eaf1d7e5178a2d116f331c340f1f7fd6de7540783ac36a7054fe9a4d64943	134	1	1	1	21
23	01/10/2009 06:06:51	e1cf3476234d8446653ad52a8939ed792003eefdcd0e897319ab9d2cb4c14c8c	134	1	1	1	22
24	01/10/2009 06:17:57	43c39b8b3728c6cb26fae0fc18803ca6cf43e15fde218e3dfbd54e633cf6753e	134	1	1	1	23
25	01/10/2009 06:24:06	9b0f52332d7d013b49016416c4818b5abb80b01ca526b7b813830348ad2321be	134	1	1	1	24
```

JD/csv

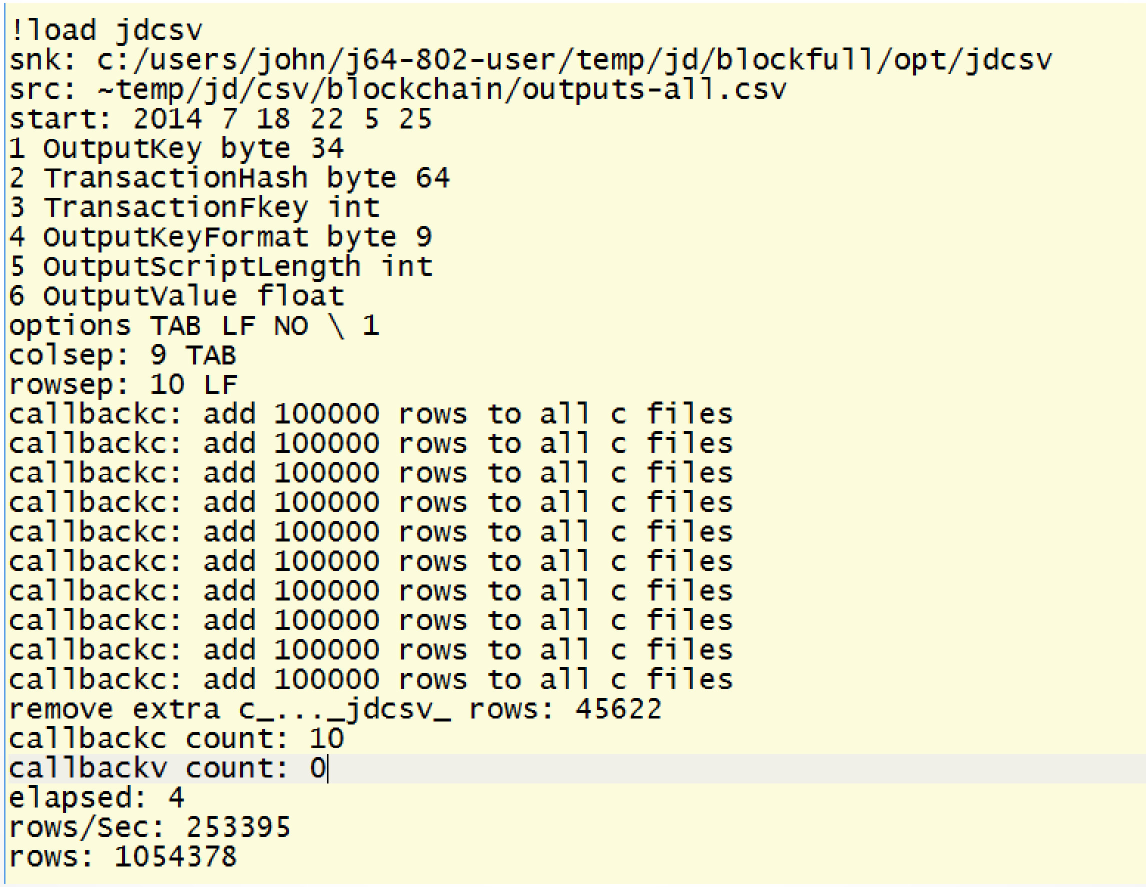
The TAB delimited transaction, input, and output files were loaded with JD's CSV import



```
NB. copy raw data to csv import
B=: (jpath '~BitJDData'),'jdcsv/'
shell (winpathsep 'xcopy ',B,'inputs2009.csv' ,' ' ,jpath F),' /s'
shell (winpathsep 'xcopy ',B,'outputs2009.csv' ,' ' ,jpath F),' /s'
shell (winpathsep 'xcopy ',B,'transactions2009.csv',' ',jpath F),' /s'

NB. copy column defs
shell (winpathsep 'xcopy ',B,'inputs2009.cdefs' ,' ' ,jpath F),' /s'
shell (winpathsep 'xcopy ',B,'outputs2009.cdefs' ,' ' ,jpath F),' /s'
shell (winpathsep 'xcopy ',B,'transactions2009.cdefs',' ',jpath F),' /s'

NB. load transactions
jd 'droptable trn'
jd 'csvrd /rows 0 transactions2009.csv trn'

NB. load inputs
jd 'droptable ipt'
jd 'csvrd /rows 0 inputs2009.csv ipt'

NB. load outputs
jd 'droptable opt'
jd 'csvrd /rows 0 outputs2009.csv opt'

NB. set references
jd 'reference trn TransactionFkey ipt TransactionFkey'
jd 'reference trn TransactionFkey opt TransactionFkey'
```

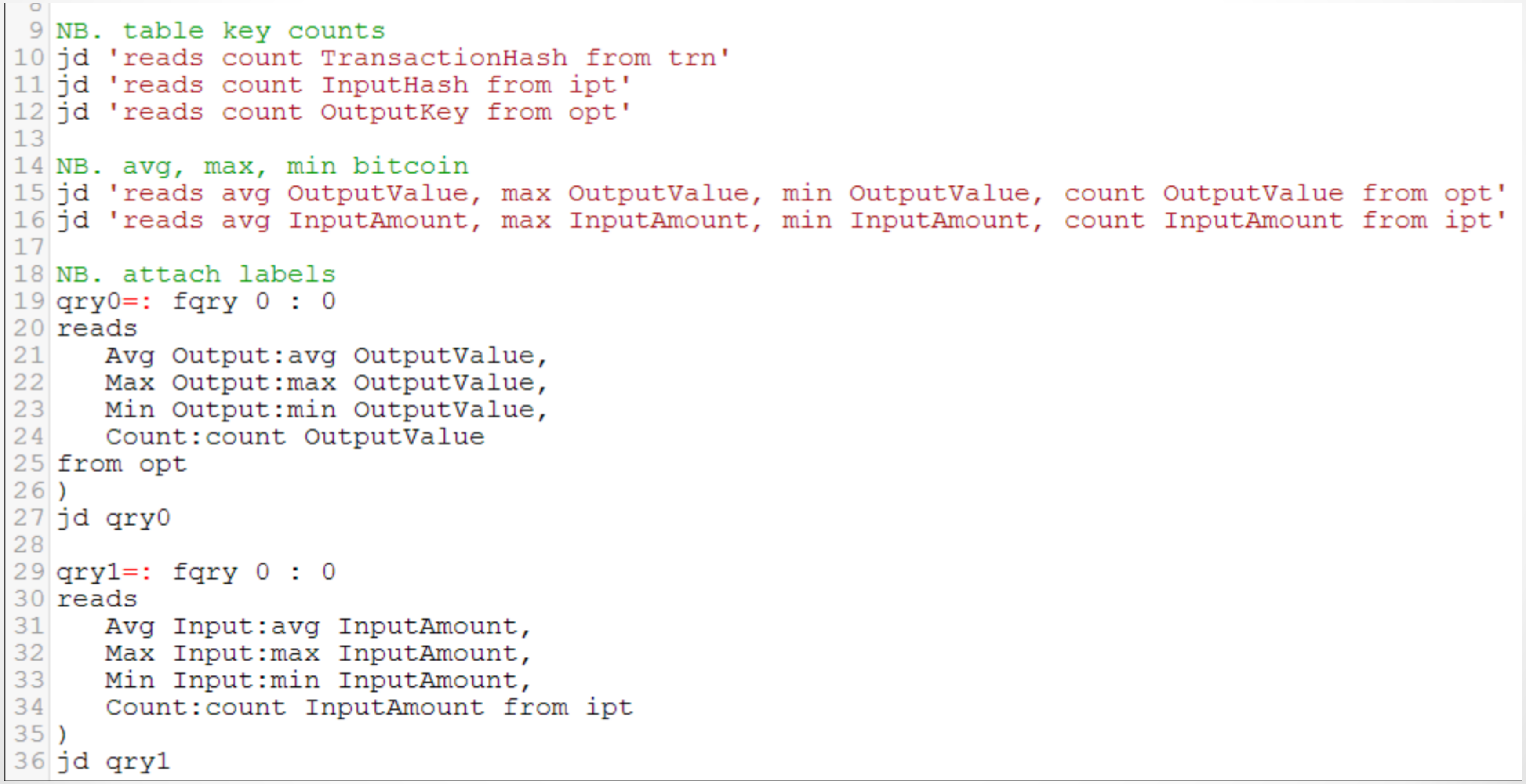
JD Load Times

- The JD CSV loader quickly loads data.



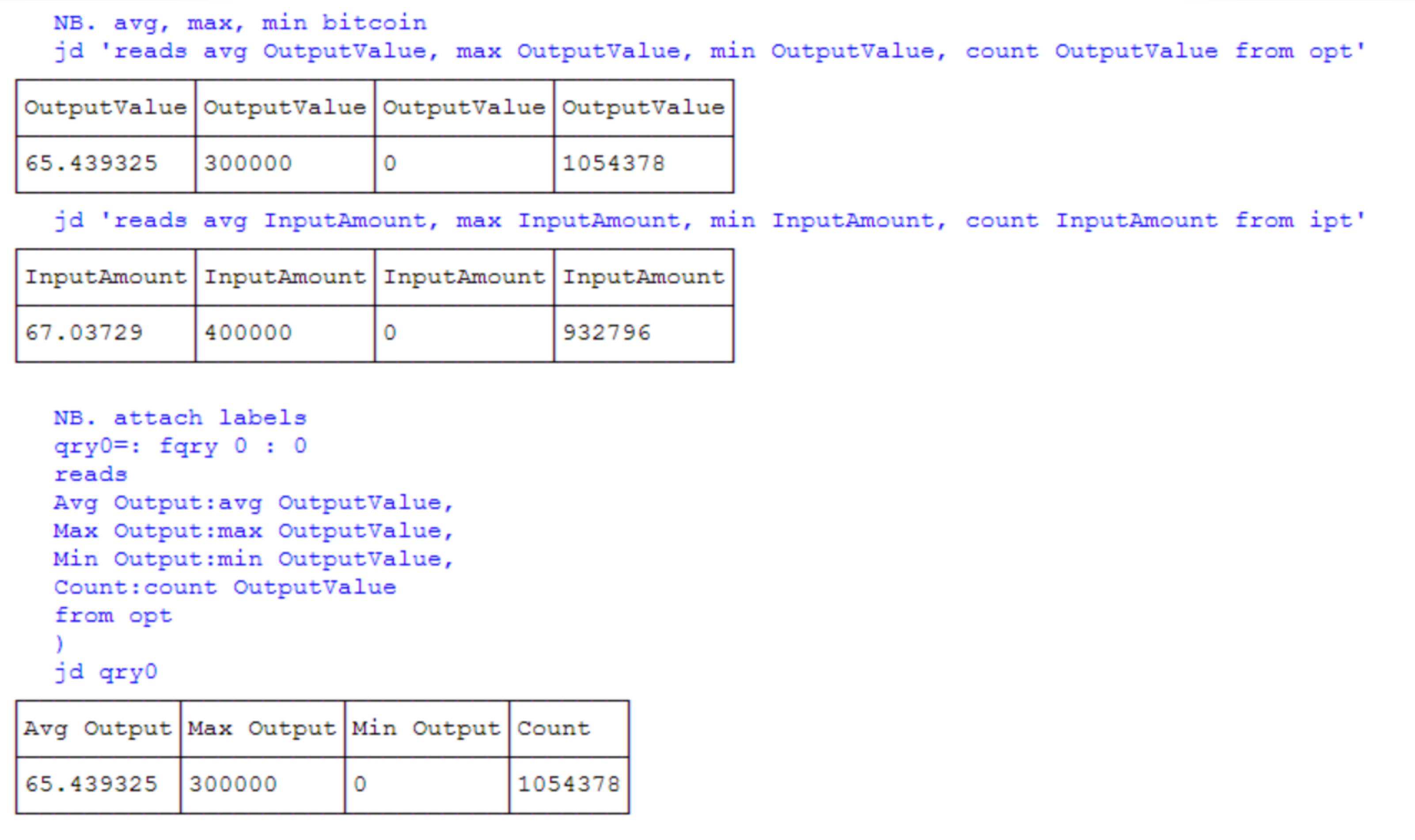
```
!load jdcsv
snk: c:/users/john/j64-802-user/temp/jd/blockfull/opt/jdcsv
src: ~temp/jd/csv/blockchain/outputs-all.csv
start: 2014 7 18 22 5 25
1 OutputKey byte 34
2 TransactionHash byte 64
3 TransactionFkey int
4 OutputKeyFormat byte 9
5 OutputScriptLength int
6 OutputValue float
options TAB LF NO \ 1
colsep: 9 TAB
rowsep: 10 LF
callbackc: add 100000 rows to all c files
callbackc: add 100000 rows to all c files
callbackc: add 100000 rows to all c files
callbackc: add 100000 rows to all c files
callbackc: add 100000 rows to all c files
callbackc: add 100000 rows to all c files
callbackc: add 100000 rows to all c files
callbackc: add 100000 rows to all c files
callbackc: add 100000 rows to all c files
callbackc: add 100000 rows to all c files
remove extra c_..._jdcsv_ rows: 45622
callbackc count: 10
callbackv count: 0
elapsed: 4
rows/Sec: 253395
rows: 1054378
```

Some JD Queries



```
NB. table key counts
jd 'reads count TransactionHash from trn'
jd 'reads count InputHash from ipt'
jd 'reads count OutputKey from opt'

NB. avg, max, min bitcoin
jd 'reads avg OutputValue, max OutputValue, min OutputValue, count OutputValue from opt'
jd 'reads avg InputAmount, max InputAmount, min InputAmount, count InputAmount from ipt'

NB. attach labels
qry0=: fqry 0 : 0
reads
   Avg Output:avg OutputValue,
   Max Output:max OutputValue,
   Min Output:min OutputValue,
   Count:count OutputValue
from opt
)
jd qry0

qry1=: fqry 0 : 0
reads
   Avg Input:avg InputAmount,
   Max Input:max InputAmount,
   Min Input:min InputAmount,
   Count:count InputAmount from ipt
)
jd qry1
```

Some JD Results



```
NB. avg, max, min bitcoin
jd 'reads avg OutputValue, max OutputValue, min OutputValue, count OutputValue from opt'
```

OutputValue	OutputValue	OutputValue	OutputValue
65.439325	300000	0	1054378

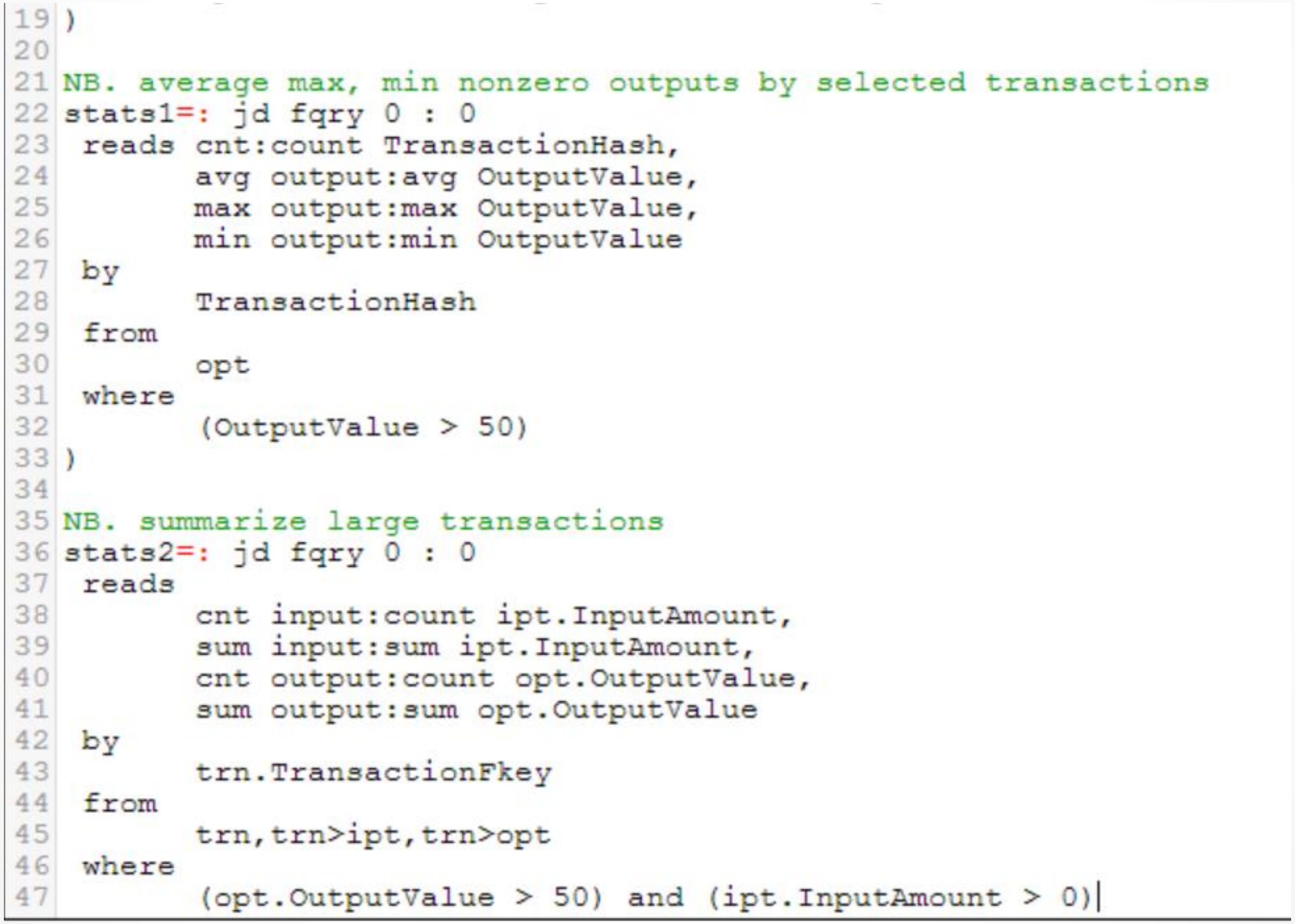
```
jd 'reads avg InputAmount, max InputAmount, min InputAmount, count InputAmount from ipt'
```

InputAmount	InputAmount	InputAmount	InputAmount
67.03729	400000	0	932796

```
NB. attach labels
qry0=: fqry 0 : 0
reads
Avg Output:avg OutputValue,
Max Output:max OutputValue,
Min Output:min OutputValue,
Count:count OutputValue
from opt
)
jd qry0
```

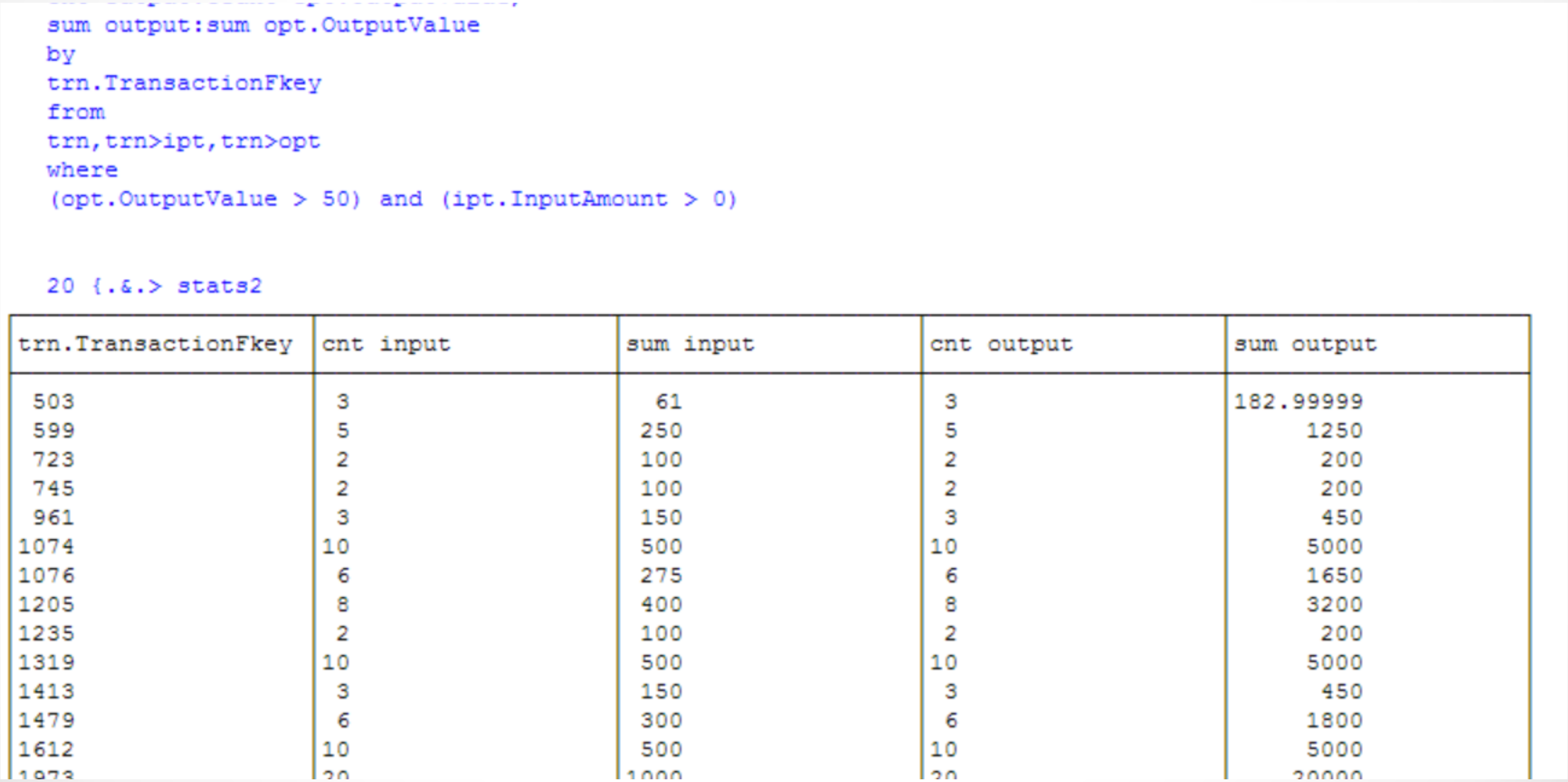
Avg Output	Max Output	Min Output	Count
65.439325	300000	0	1054378

More JD Queries



```
)

NB. average max, min nonzero outputs by selected transactions
stats1=: jd fqry 0 : 0
 reads cnt:count TransactionHash,
       avg output:avg OutputValue,
       max output:max OutputValue,
       min output:min OutputValue
 by
       TransactionHash
 from
       opt
 where
       (OutputValue > 50)
)

NB. summarize large transactions
stats2=: jd fqry 0 : 0
 reads
       cnt input:count ipt.InputAmount,
       sum input:sum ipt.InputAmount,
       cnt output:count opt.OutputValue,
       sum output:sum opt.OutputValue
 by
       trn.TransactionFkey
 from
       trn,trn>ipt,trn>opt
 where
       (opt.OutputValue > 50) and (ipt.InputAmount > 0)
```

More JD Results

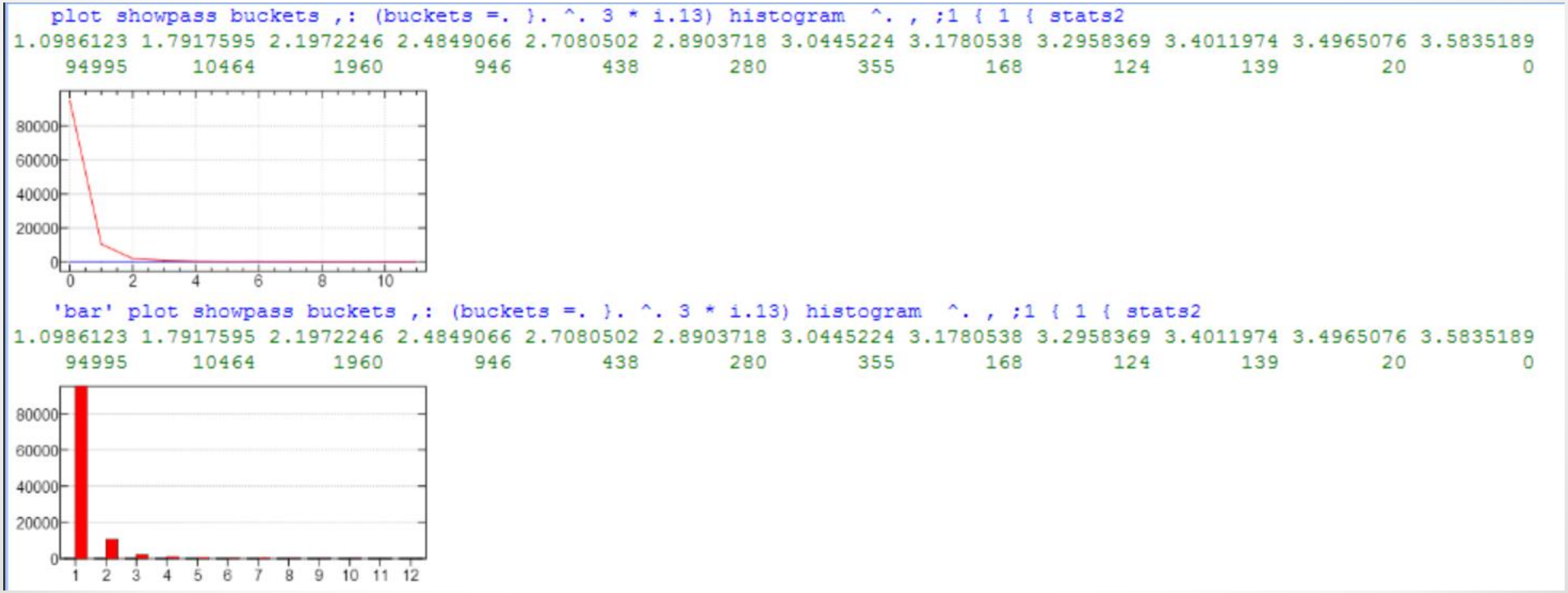


```
sum output:sum opt.OutputValue
by
trn.TransactionFkey
from
trn,trn>ipt,trn>opt
where
(opt.OutputValue > 50) and (ipt.InputAmount > 0)


20 {.&.> stats2
```

trn.TransactionFkey	cnt input	sum input	cnt output	sum output
503	3	61	3	182.99999
599	5	250	5	1250
723	2	100	2	200
745	2	100	2	200
961	3	150	3	450
1074	10	500	10	5000
1076	6	275	6	1650
1205	8	400	8	3200
1235	2	100	2	200
1319	10	500	10	5000
1413	3	150	3	450
1479	6	300	6	1800
1612	10	500	10	5000
1973	20	1000	20	20000

Blockchain Spelunking

- JD query results are J nouns that can be easily analyzed within J.



```
   plot showpass buckets ,: (buckets =. }. ^. 3 * i.13) histogram  ^. , ;1 { 1 { stats2
1.0986123 1.7917595 2.1972246 2.4849066 2.7080502 2.8903718 3.0445224 3.1780538 3.2958369 3.4011974 3.4965076 3.5835189
    94995     10464      1960       946       438       280       355       168       124       139        20         0
```

```
   'bar' plot showpass buckets ,: (buckets =. }. ^. 3 * i.13) histogram  ^. , ;1 { 1 { stats2
1.0986123 1.7917595 2.1972246 2.4849066 2.7080502 2.8903718 3.0445224 3.1780538 3.2958369 3.4011974 3.4965076 3.5835189
    94995     10464      1960       946       438       280       355       168       124       139        20         0
```

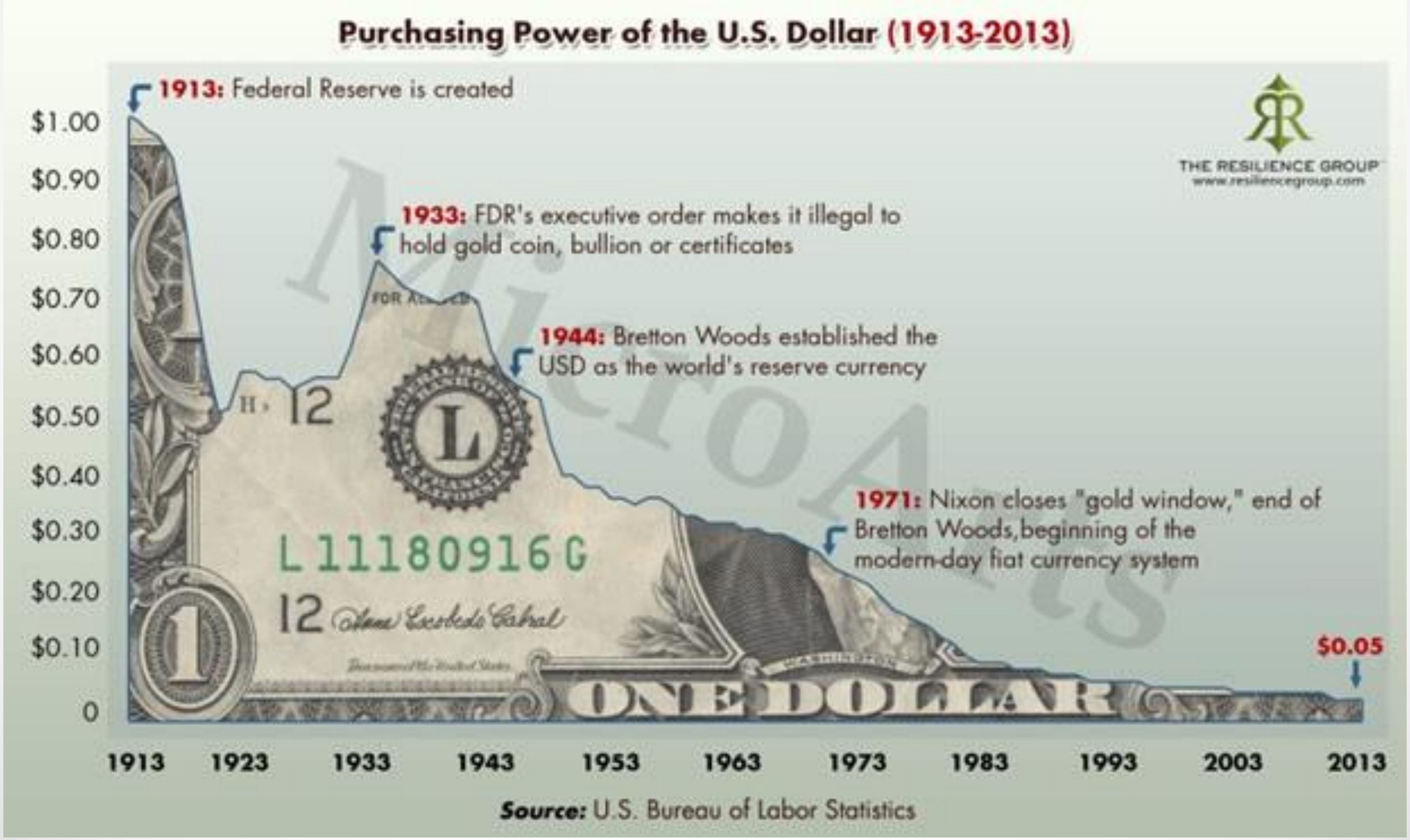
JD Impressions

1. J luxury compared to ODBC,ADO,SQL Server, Oracle et cetera.
2. Having data in native J formats greatly simplifies spelunking.
3. CSV load/unload fast and effective.
4. Requires a reasonable level of J expertise.
5. Use of native files makes backups, copies, and other maintenance tasks dead easy.
6. Probably not for the paranoid.
7. Needs efficient front ends (C#, Java) to plug into corporate data systems.

Bitcoin Impressions

1. Simple and direct software.
2. Elegant underlying thesis.
3. The first original idea about money in decades.
4. The solution of the "double spend problem" has implications far beyond cash exchanges.
5. Bitcoin does not provide adequate cover for criminals and tax cheats.
6. The scale of the current system (2014) is tiny compared to alternatives.
7. Already more secure and counterfeit proof than any national currency.

Bitcoin ≡ Fiat Doom



Purchasing Power of the U.S. Dollar (1913-2013)

1913: Federal Reserve is created

1933: FDR's executive order makes it illegal to hold gold coin, bullion or certificates

1944: Bretton Woods established the USD as the world's reserve currency

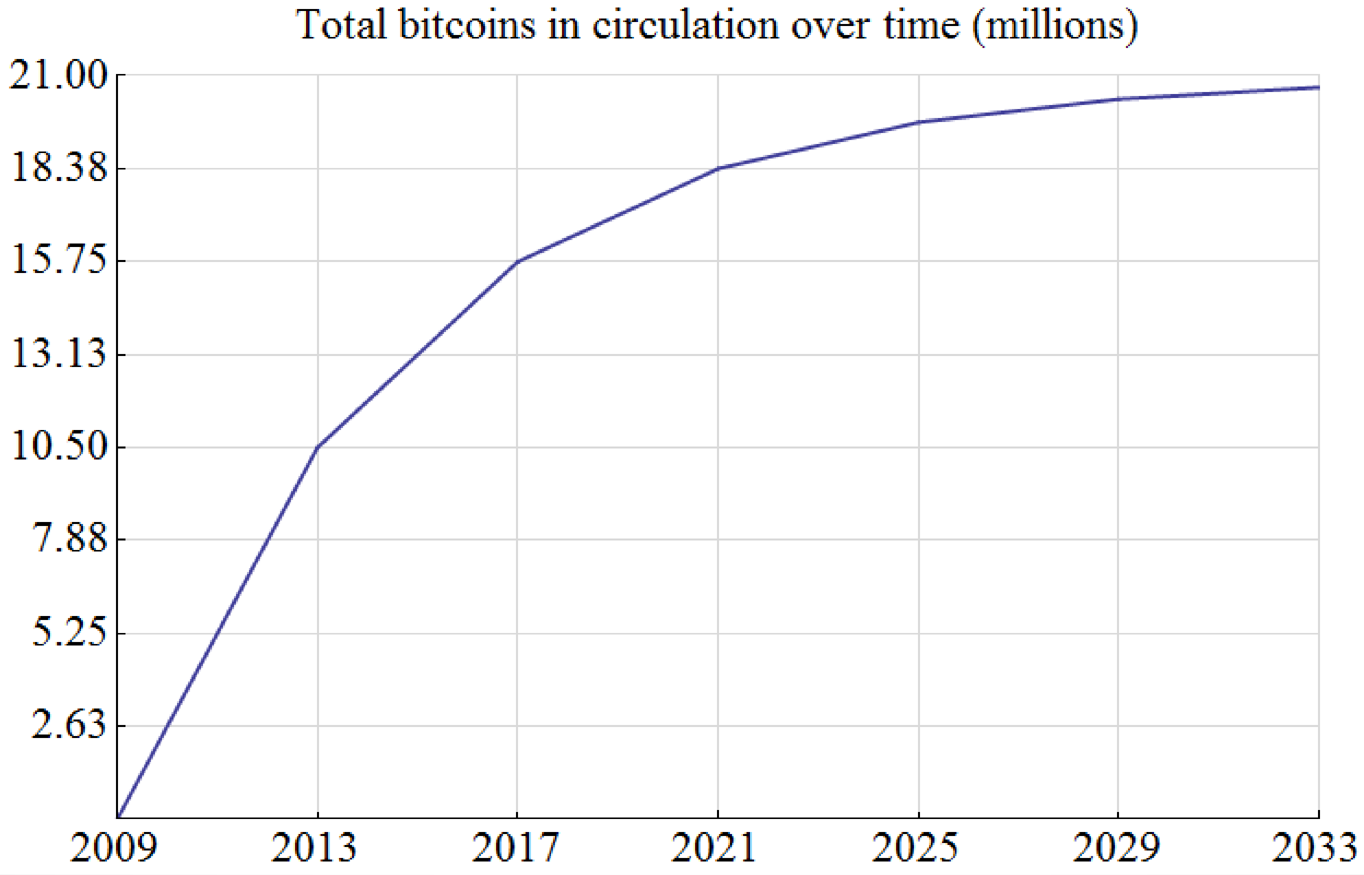
1971: Nixon closes "gold window," end of Bretton Woods,beginning of the modern-day fiat currency system

$0.05

Source: U.S. Bureau of Labor Statistics

Bitcoin ≡ Fiat Doom

- What will be the FED inflation rate five years from now?



For the Bitcoin Curious

1. The main website:
   https://bitcoin.org/en/
2. Satoshi’s paper:
   https://bitcoin.org/bitcoin.pdf
3. Be a Bitcoin node:
   https://bitcoin.org/en/download
4. Developer’s:
   https://bitcoin.org/en/bitcoin-for-developers