Here are the results that were missing from the Bitcoin presentation.

The first step was to process the input data in a more efficient way without using Hadoop. The CSV file was sorted by key (bitcoin account ID) using GnuUtils sort (called external sorting). It took about 30 minutes for a 13 GB CSV.

Sample Data (~150 million transactions):

bitcoin address (account id)	timestamp	day	bitcoins transferred	$ price per bitcoin
70a3f2149e56e0351f4b0328368bafa7b1e0785a	1398178160	2014-04-22	-0.01200546	494.54
70a3f2149e56e0351f4b0328368bafa7b1e0785a	1398178160	2014-04-22	-0.01507263	494.54
71c9930a8cb467b0416447f923a0eaf035aeceb3	1398178160	2014-04-22	0.01010339	494.54
76f7b17312424fcee33e05421be3b7a6ad0d4291	1398178160	2014-04-22	0.02239569	494.54

sort.exe -t -k 1 -T s:/temp/ s:/bitcoin-transactions-04-22-2014.csv -o s:/sorted.csv

Sample Data (~150 million transactions):

bitcoin address (account id)	timestamp	day	bitcoins transferred	$ price per bitcoin
00b984005d1f23c6590ac6b8c081702e782ab773	1384979745	2013-11-20	35.09674580	542.03
00b984005d1f23c6590ac6b8c081702e782ab773	1384987133	2013-11-20	-35.09674580	542.03
00b98462c53274af3bb887fccd6643363ce4a6140	1362080097	2013-02-28	0.67919408	33.10
00b98462c53274af3bb887fccd6643363ce4a6140	1362086726	2013-02-28	-0.67919408	33.10

A Java 1.7 program was used to calculate the profit and tax on the 13 GB sorted CSV file. Running time is about 150 seconds and the output is 334 MB.

java OptimizedBitcoinTaxCalc s:/sorted.csv s:/complex-results.csv

Sample data (~4.7 million taxable accounts):

bitcoin address (account id)	short term profit	short term tax	long term profit	long term tax	profit minus tax
000192e5c23d88f4c55f b9ebbd5bbfa617fac9d1	$3.36	$1.18	$0.00	$0.00	$2.19
000194de5ee67dca0f7d 1322168ba1a70b9b99cf	$1125.90	$394.06	$0.00	$0.00	$731.83
00019568303c6be166b1 3b06f82a7af8429c01f3	$2.75	$0.96	$0.00	$0.00	$1.79
000199451851650c8850 733a71756a90f15d3d8a	$266.60	$93.31	$0.00	$0.00	$173.29

A different Java 1.7 program was used to sum the profit and tax for all bitcoin addresses.

java BitcoinAggregator s:/complex-results.csv

shortTermProfit: $3,418,830,016.23 ($3.4 billion)
shortTermTax: $1,196,590,566.78 ($1.2 billion)
longTermProfit: $489,018,541.45 ($489 million)
longTermTax: $73,352,797.84 ($73 million)
totalProfitMinusTax: $2,637,905,476.49 ($2.6 billion)

totalTax: $1,269,943,364.62 ($1.27 billion)
Took: 18 seconds

Totals
Input Data: 13 GB csv, 150 million transactions

Conclusion
Most people own bitcoin for less than a year. The results seem realistic because there are currently about 12 million bitcoins in existence and the price for 1 bitcoin is about $400.00. The market value of all bitcoins is around $5 billion.

We calculated the tax rate as though all transactions occurred in the United States however tax rate [varies by country](). Also, not every transfer from one bitcoin address to another is a buy/sell operation. A user may be transferring bitcoins between their different accounts. A user can create many accounts as they like for free.

For security purposes, some users only use a bitcoin addresses once. This makes it more difficult to track the transaction graph. Since Bitcoin cannot be linked directly to users and their country, it's difficult to tax.

One source of data about profits made is Bitcoin exchanges that convert bitcoins into dollars. These companies are likely to provide information to taxation bureaus.

References
Files mentioned can be found in: s3://group-haifa-amble-jesse/

https://www.gainskeeper.com/siebert/siebert_fund.htm