**完整性：**

数据总体非常完整，没有明显的中断，也没有发现明显的缺失。

数据总共145万行左右。

非完整行数（即dt.ncomplete <- filter(dt, !complete.cases(dt)) ）有60万行。

具体检查了每秒是否都有数据，发现仅一处遗失数据（发生在11:11:12秒，但应该是该秒就是没有数据传输的）。

此外随机找了一个合约画图，没有发现明显问题。

**分合约数据：**

分合约的数据行数在contract.length.csv文件内。

另外检查了哪些数据是区分500 ms的，这个记录在500 ms.csv内。应该如之前设想的上期所的才分500 ms。

**缺失数据处理：**

对非完整的60万行进行了审查。按照有没有500ms数据（等效于是否是上期所合约）检查了这些行的NA值发生在哪些变量。

对于上期所数据，NA值发生在V9-V13和V19-V23这些变量。其中V9-V13为AskPrice1到AskPrice5；V19-V23为BidPrice1到BidPrice5。另外一般的缺失情形是AskPrice和BidPrice的2到4档同时缺失；少数情况AskPrice1或者BidPrice1连同两者的2到4档同时缺失（如果一档价格缺失，两者的2-4档价格就都是缺失的）。不存在其他情况。

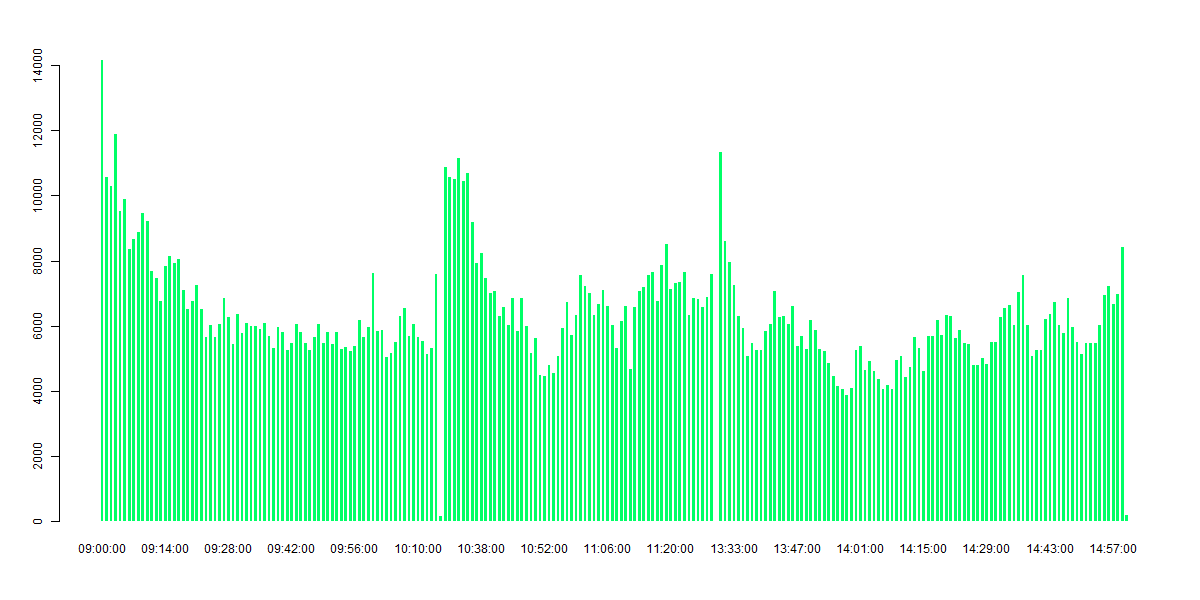
对于非上期所的数据，NA值都发生在一档价格，两个一档价格是否缺失没有关联性。其他变量都不存在缺失。

**流量平稳性：**

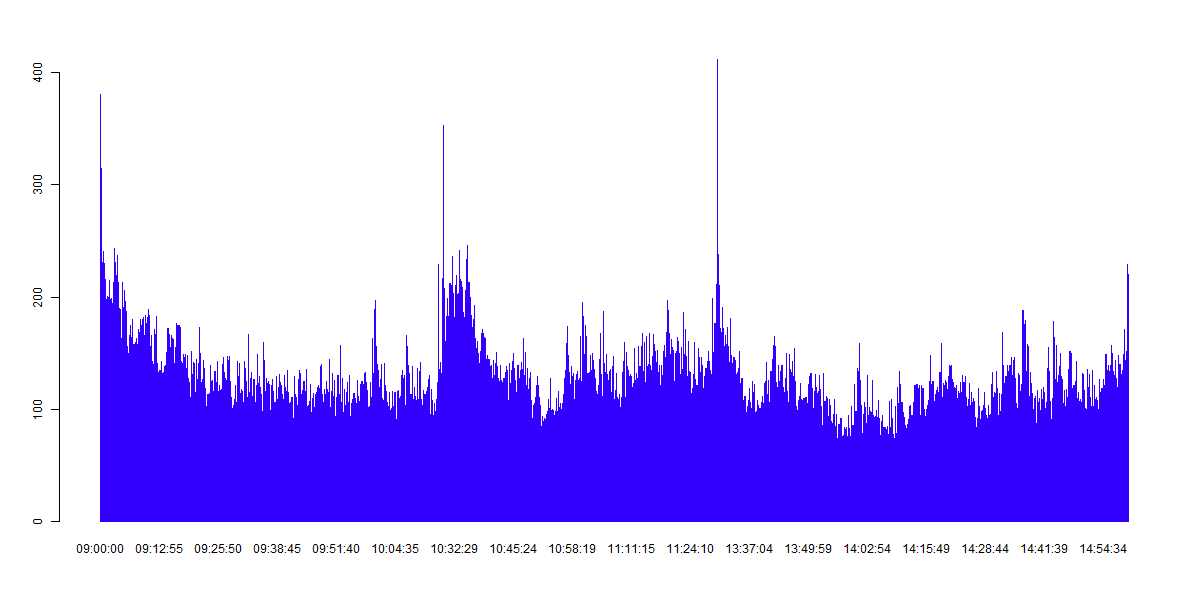
这里估算了数据流入的平稳性和流入流量。

平均每分钟流入数据6350行，中位数6061行。最大一分钟流入14200行。

秒流量（在文件HM.volume.csv）：



平均每秒流入107行数据，中位数102行。最大一秒流入412行（在文件HMS.volume.csv）:



大部分时间流量较为平稳。9:00:00开盘后5分钟，10:35:00前后流入数据较多（早盘休息过后），这些时间的流量大概在每秒200行，是平稳流量的两倍；以及下午13:00:00开盘瞬时有一个高流量，大概400行。