In 1970, I was just expelled from the University, and I was looking for something interesting to do, when I was introduced to Pista. Pista thought (wrongly) that because I was physics student I must be able to do programming and I could do the data processing. When I told him that this is not the case, he said don't worry you will learn, and he was right about this. He also asked whether I wanted to participate in the fieldwork, and we repeated the same conversations, I told that I haven't got the foggiest, and he said don't worry. In this he was not quite as right, although I did learn a lot about sociology - slowly, I did not become an expert immediately. I did do interviews (in-depth and questionnaire) in both the Gypsy and Poverty surveys. The Poverty survey was processed in the KSH and I did the Gypsies.Gypsy questionnaire has a complex structure, you might not have noticed that, because today there are standard software tools which can handle such data, but in the 70's these tools were not invented yet. We were just one step forward of the Hollerith machines. People still did the tabulations by mechanically sorting cards (although not for long) and even for those who have already moved on to computers the legacy of the Hollerith machines lived on for years. The sample and the questionnaire were both household based, but we identified families withing a household and individuals with a family (adults and children). Today processing this kind of data is a trivial task in any relational data base, but at the time SQL was not invented yet. But I am getting ahead of things. The data was punched on cards, which had to be read into the MTA mainframe (CDC3300). If I remember correctly we had some 26 boxes of punch-cards (2000 cards in a box). That was a lot of cards and it took several hours to read them all (the card reader often jammed and we had to re-punch the damaged card before we could continue). The problem was that we were allocated only 10 min of computer time for the whole project vis-a-vis several hours for just reading the cards. I figured out a system how to fool the computer's accounting mechanism to measure only the CPU time and not the actual block time.The data was then stored on a magnetic tape (MT). The reason for that was that neither FORTRAN nor the operating system (MASTER) was able to handle variable length records on a disk. By using MT I was able to trick them. It was a lot of cards, but not a lot of data (not even in those days' terms), but there was a catch. The MT unit puts a .75 inch gap between records and we had over 50,000 cards. If I kept them as separate records (as I was reading them) it wouldn't have fit on one 2400 feet tape. I had to compact the data into long but variable length household records, which, at the processing time, I had to parse and break up into its hierarchical components.Pista wanted 3 dimensional tables, should SPSS or SAS were available it would have been a trivial task, but we could not wait several years until they reached Hungary. Although (even in those days) FORTRAN could easily handle 3 dimensional arrays but reading (comprehending) such tables can be very confusing unless they are well labeled. I've produced many kilos, stacks and stacks of tables. Printing meaningful labels on the tables was unusual at the time. Again in those day there were no tools for handling metadata. It was a lot of work to punch the labels on cards and make sure that the labels match the actual data. I could only hope, but couldn't be sure that I got is right all the time.The funny thing is that, at the time, I had no idea how difficult was what I've done, I only realized it decades later. Survey data processing was a very small field in the early 70's, only a handful of people were doing it and we all knew each other. I looked at them with sort of embarrassment that they were able to produce tables at a much faster rate than me, and with very little effort. I understood much later that they were using simple data structures (flat files with fix record length) and codes of only 1-9. They printed the codes not actual labels as heading, and the tables were also only numbered and not labeled.Now a few words about what we (I) didn't do, which could have been or perhaps should have been done.First of all, I did not do any significance test. It was not requested from me, and

did not occur to me either, at all. If a figure was larger that another figure we took it at face value. The significance tests were not really necessary, we were looking for major effects which were "obviously" significant. The second one was more important. One of the objective of the study was to estimate the number of Gypsies (ie persons). But we had a household based sample, with varying number persons living in each household. Estimating the number of persons from a household sample is not a trivial task. I was aware of the problem but did not have a solution and therefore I did not raise it with Pista. Again, in reality this was not such a big issue, since even this very rough estimate was extremely valuable, and a more precise estimate wouldn't have made much of a difference in practice.What happened to the tapes later, I don't know. To my knowledge there was no budget (or perhaps not even interest) to archive the data. The tapes were rented from the Computer Center, and I assume if nobody paid for the rent the tapes were simply reused, overwritten. In later projects that I was involved with, I made backups, and actually stole the backup tapes and took them home. I don't remember whether I've done the same with this data set. It was my first such project and probably wasn't smart enough yet. This kind of backup was only good for short periods anyway, as magnetic tapes need to be refreshed (by copying) regularly. Finally the credits. I sincerely owe a lot to the Computer Center's staff. The software engineers spent infinite time teaching us how to write code and how to debug the programs, the Center's manager (Olah Eva) was particularly helpful allowing me access to the computer room at night, thus enabling me to finish the work on time, and who can forget the punch-girls, all of whom I was (hopelessly) in love with.