

Wrangler
Worldcup.txt

Q: Determine which Production/Distribution house produced maximum movies.

***Universal Studios (Universal Pictures)** has the most films however, Warner Bros. will have more if we consider the smaller companies like MGM and New Line Cinema are considered as Warner Bros since they are owned by it. I believe they are separate studios so Universal has more moviess.*

worldcup.txt

Q. According to the dataset, how often has each country won the world cup?

ARG	2
BRA	5
ENG	1
ESP	1
FRA	1
GER	3
ITA	4
URU	2

Discussion Questions

Q: From your experience, briefly discuss the pro and cons between using Data Wrangler as compared to lower levels tools like sed/awk?

I found the Data Wrangler to be an excellent tool that is very powerful and something that I expect to use in the future. It is excellent in that it visualizes the transformations one can do and makes the job of cleaning up messy data simpler. In addition, the pivoting and fill operations have been very crucial in cleaning up the worldcup.txt data, excelling beyond awk and sed. However, it is a bit buggy and appears unable to handle large data which the lower level tools seem to be better at. Sed and Awk are excellent tools that I have used before. They excel in the fact that they allow so much flexibility in munging data in any way one likes. I found pivoting to be a bit harder in these tools than the wrangler.

Q: What additional operations would have made using Data Wrangler "easier"?

The wrangler has a lot of good operations and is very powerful. What would make it best is if it had a script tab where one can just write low level scripts in the language used without having to use the graphic part.