Data cleaning principles

fundamentals

- 1. Don't clean data when tired or hungry
- 2. Don't trust anyone (even yourself)
- 3. Think about what might have gone wrong
- 4. Use care in merging
- 5. Dates & categories suck

verify

- 6. Verify that distinct things are distinct
- 7. Verify that matching things match
- 8. Check calculations
- 9. Look for other instances of problems

explore

- 10. Make lots of plots
- 11. Look at missing value patterns
- 12. With big data make more plots
- 13. Follow up all artifacts

ask

- 14. Ask questions
- 15. Ask for the primary data
- 16. Ask for metadata
- 17. Ask why data are missing

document

- 18. Create checklists & pipelines
- 19. Document not just what but why
- 20. Expect to recheck