Database Design
WV Senate Problem

Topics

- Determine if Excel or Access are best suited for a dataset
- Design a relational database including tables and fields
- Identify appropriate table relationships

Background Information

This project includes information on West Virginia Senate elections from 2000 to 2014.

Instructions

IMPORTANT: Complete the steps below in the order they are given. Completing the steps out of order may complicate the assignment or result in an incorrect result.

NOTE: This project will be done entirely in Microsoft Excel. This project involves designing a database, not creating one. Database creation is covered in a separate project.

1. Download and extract the provided Data Files ZIP file. It contains the following files for use in this assignment:

a. **senate_ppdd_wvsp.xlsx** – Excel spreadsheet with information on West Virginia Senate elections from 2000-2014 [1]–[5].

Column Name	Type	Description
Year	Number	Year of election.
District	Number	State Senate district number.
Counties	Text	List of counties or parts of counties in the district. Counties partially within the district are indicated by (P).
Population	Number	Population of the district.
Redistricting Cycle	Number	Census data year used for redistricting.
Incumbent	Text	Name of the incumbent.
Incumbent Retired	Yes/No	Indication if incumbent was not running.
Winner	Text	Name of the winner of the election.
Winner Pctg	Number	Percentage of the total vote received by the winner.
American Freedom Candidate	Text	Name of the American Freedom Party candidate.
American Freedom Raised	Currency	Amount of funds raised by the candidate.
American Freedom Votes	Number	Votes received by the candidate.
Constitution Candidate	Text	Name of the Constitution Party candidate.
Constitution Raised	Currency	Amount of funds raised by the candidate.
Constitution Votes	Number	Votes received by the candidate.
Democratic Candidate	Text	Name of the Democratic Party candidate.
Democratic Raised	Currency	Amount of funds raised by the candidate.
Democratic Votes	Number	Votes received by the candidate.
Libertarian Candidate	Text	Name of the Libertarian Party candidate.
Libertarian Raised	Currency	Amount of funds raised by the candidate.
Libertarian Votes	Number	Votes received by the candidate.
Mountain Candidate	Text	Name of the Mountain Party candidate.
Mountain Raised	Currency	Amount of funds raised by the candidate.
Mountain Votes	Number	Votes received by the candidate.
Natural Law Candidate	Text	Name of the Natural Law Party candidate.
Natural Law Raised	Currency	Amount of funds raised by the candidate.
Natural Law Votes	Number	Votes received by the candidate.
Republican Candidate	Text	Name of the Republican Party candidate.
Republican Raised	Currency	Amount of funds raised by the candidate.
Republican Votes	Number	Votes received by the candidate.
Write-In Candidate	Text	Name of the Write-In candidate.
Write-In Raised	Currency	Amount of funds raised by the candidate.
Write-In Votes	Number	Votes received by the candidate.

2. Open the **senate_ppdd_wvsp.xlsx** file to view and understand the data it contains.

Determine if Excel or Access are best suited for a dataset

3. We need to determine if this dataset is better suited for a Microsoft Excel workbook or a Microsoft Access database. Access databases are more complex, but they can offer some advantages in organizing larger datasets.

    a. Consider the following questions:

        i.   Are there multiple categories of data, and if so, is there a relationship between the different categories? If there are, the data can be split into multiple tables in an Access database and related together. If there are not, Excel might be a simpler way to organize the data.

        ii. Are there large amounts of redundant data? If there are, Access may be better because its relational design can reduce redundancy and the chance of errors. If there is not, Excel may be better because it is simpler.

        iii. Is there a need to be able to analyze only specific parts of the data? If there is, Access queries can be used to select a subset of the data. While filtering can be done in Excel, it is more cumbersome and relatively difficult to reuse search filters.

        iv. Do charts need to be created to represent the data? If charts are needed, they can only be created in Excel. When Access is used, the data first must be exported to Excel before a chart can be made.

        v.   Will what-if analysis need to be performed? Excel includes features such as trendlines, Goal Seek, and scenarios to facilitate this process. Access is better suited to working on existing data and does not include built-in tools for what-if analysis.

    b. Look at your answers from Step 3a. In many cases, there will be a clear pattern showing if Excel or Access should be used. If there is not, you may need to store your data in an Access database and export portions of it to Excel for detailed analysis.

        For this particular dataset, Access is best suited to store the information. We have multiple categories of data, much of it redundant, we want to be able to analyze specific parts of it, and we don't need to create charts or perform what-if analysis.

Database Design
WV Senate Problem

Identify field groupings

4. Review the types of information (fields/columns) that need to be stored in the database. Create groups based on related types of information.

    In this dataset, the information broadly belongs to two categories: district information and election information.

Column Name	Field Grouping
Year	Elections
District	Districts
Counties	Districts
Population	Districts
Redistricting Cycle	Districts
Incumbent	Elections
Incumbent Retired	Elections
Winner	Elections
Winner Pctg	Elections
American Freedom Candidate	Elections
American Freedom Raised	Elections
American Freedom Votes	Elections
Constitution Candidate	Elections
Constitution Raised	Elections
Constitution Votes	Elections
Democratic Candidate	Elections
Democratic Raised	Elections
Democratic Votes	Elections
Libertarian Candidate	Elections
Libertarian Raised	Elections
Libertarian Votes	Elections
Mountain Candidate	Elections
Mountain Raised	Elections
Mountain Votes	Elections
Natural Law Candidate	Elections
Natural Law Raised	Elections
Natural Law Votes	Elections
Republican Candidate	Elections
Republican Raised	Elections
Republican Votes	Elections
Write-In Candidate	Elections
Write-In Raised	Elections
Write-In Votes	Elections

Database Design
WV Senate Problem

5. Look at the fields and groupings you have identified. See if they can be broken down further.

We can create a subgroup for election candidates.

Fields with changed groupings are highlighted in yellow below:

Column Name	Field Grouping
Year	Elections
District	Districts
Counties	Districts
Population	Districts
Redistricting Cycle	Districts
Incumbent	Elections
Incumbent Retired	Elections
Winner	Elections
Winner Pctg	Elections
American Freedom Candidate	Elections – Candidates
American Freedom Raised	Elections – Candidates
American Freedom Votes	Elections – Candidates
Constitution Candidate	Elections – Candidates
Constitution Raised	Elections – Candidates
Constitution Votes	Elections – Candidates
Democratic Candidate	Elections – Candidates
Democratic Raised	Elections – Candidates
Democratic Votes	Elections – Candidates
Libertarian Candidate	Elections – Candidates
Libertarian Raised	Elections – Candidates
Libertarian Votes	Elections – Candidates
Mountain Candidate	Elections – Candidates
Mountain Raised	Elections – Candidates
Mountain Votes	Elections – Candidates
Natural Law Candidate	Elections – Candidates
Natural Law Raised	Elections – Candidates
Natural Law Votes	Elections – Candidates
Republican Candidate	Elections – Candidates
Republican Raised	Elections – Candidates
Republican Votes	Elections – Candidates
Write-In Candidate	Elections – Candidates
Write-In Raised	Elections – Candidates
Write-In Votes	Elections – Candidates

Reorganize the fields

6. Review the field list. Are there fields that contain multiple distinct pieces of information that can be split into multiple fields?

   In this example, the names for the incumbent, winner, and each candidate can be split into separate fields for the first name and the last name.

   Changed fields are highlighted in yellow below:

Column Name	Field Grouping
Year	Elections
District	Districts
Counties	Districts
Population	Districts
Redistricting Cycle	Districts
Incumbent First Name	Elections
Incumbent Last Name	Elections
Incumbent Retired	Elections
Winner First Name	Elections
Winner Last Name	Elections
Winner Pctg	Elections
American Freedom Candidate First Name	Elections – Candidates
American Freedom Candidate Last Name	Elections – Candidates
American Freedom Raised	Elections – Candidates
American Freedom Votes	Elections – Candidates
Constitution Candidate First Name	Elections – Candidates
Constitution Candidate Last Name	Elections – Candidates
Constitution Raised	Elections – Candidates
Constitution Votes	Elections – Candidates
Democratic Candidate First Name	Elections – Candidates
Democratic Candidate Last Name	Elections – Candidates
Democratic Raised	Elections – Candidates
Democratic Votes	Elections – Candidates
Libertarian Candidate First Name	Elections – Candidates
Libertarian Candidate Last Name	Elections – Candidates
Libertarian Raised	Elections – Candidates
Libertarian Votes	Elections – Candidates
Mountain Candidate First Name	Elections – Candidates
Mountain Candidate Last Name	Elections – Candidates
Mountain Raised	Elections – Candidates
Mountain Votes	Elections – Candidates

Natural Law Candidate First Name	Elections – Candidates
Natural Law Candidate Last Name	Elections – Candidates
Natural Law Raised	Elections – Candidates
Natural Law Votes	Elections – Candidates
Republican Candidate First Name	Elections – Candidates
Republican Candidate Last Name	Elections – Candidates
Republican Raised	Elections – Candidates
Republican Votes	Elections – Candidates
Write-In Candidate First Name	Elections – Candidates
Write-In Candidate Last Name	Elections – Candidates
Write-In Raised	Elections – Candidates
Write-In Votes	Elections – Candidates

Divide fields into tables / Determine appropriate field types

7. Divide the field groupings into a series of tables. Identify appropriate data types and names for each field.

Table: *Districts*		
Field Name	**Type**	**Description**
RedistrictingCycle	Number – Integer	Census data year used for redistricting.
District	Number – Integer	Number of State Senate district.
Counties	Short Text	List of counties or parts of counties in the district.
Population	Number – Integer	Population of district.

Table: *Elections*		
Field Name	**Type**	**Description**
Year	Number – Integer	Year of election.
IncumbentFirstName	Short Text	First name of incumbent.
IncumbentLastName	Short Text	Last name of incumbent.
IncumbentRetired	Yes/No	Indication if the incumbent was not running.
WinnerFirstName	Short Text	First name of the winner.
WinnerLastName	Short Text	Last name of the winner.
WinnerPctg	Number – Percentage	Percentage of the total vote received by the winner.

Database Design
WV Senate Problem

Table: *Candidates*		
Field Name	**Type**	**Description**
AmericanFreedomCandidateFirstName	Short Text	First name of the candidate.
AmericanFreedomCandidateLastName	Short Text	Last name of the candidate.
AmericanFreedomRaised	Currency	Amount of funds raised by the candidate.
AmericanFreedomVotes	Number – Integer	Votes received by the candidate.
ConstitutionCandidateFirstName	Short Text	First name of the candidate.
ConstitutionCandidateLastName	Short Text	Last name of the candidate.
ConstitutionRaised	Currency	Amount of funds raised by the candidate.
ConstitutionVotes	Number – Integer	Votes received by the candidate.
DemocraticCandidateFirstName	Short Text	First name of the candidate.
DemocraticCandidateLastName	Short Text	Last name of the candidate.
DemocraticRaised	Currency	Amount of funds raised by the candidate.
DemocraticVotes	Number – Integer	Votes received by the candidate.
LibertarianCandidateFirstName	Short Text	First name of the candidate.
LibertarianCandidateLastName	Short Text	Last name of the candidate.
LibertarianRaised	Currency	Amount of funds raised by the candidate.
LibertarianVotes	Number – Integer	Votes received by the candidate.
MountainCandidateFirstName	Short Text	First name of the candidate.
MountainCandidateLastName	Short Text	Last name of the candidate.
MountainRaised	Currency	Amount of funds raised by the candidate.
MountainVotes	Number – Integer	Votes received by the candidate.
NaturalLawCandidateFirstName	Short Text	First name of the candidate.
NaturalLawCandidateLastName	Short Text	Last name of the candidate.
NaturalLawRaised	Currency	Amount of funds raised by the candidate.
NaturalLawVotes	Number – Integer	Votes received by the candidate.
RepublicanCandidateFirstName	Short Text	First name of the candidate.

Database Design
WV Senate Problem

Table: *Candidates*		
Field Name	**Type**	**Description**
RepublicanCandidateLastName	Short Text	Last name of the candidate.
RepublicanRaised	Currency	Amount of funds raised by the candidate.
RepublicanVotes	Number – Integer	Votes received by the candidate.
WriteInCandidateFirstName	Short Text	First name of the candidate.
WriteInCandidateLastName	Short Text	Last name of the candidate.
WriteInRaised	Currency	Amount of funds raised by the candidate.
WriteInVotes	Number – Integer	Votes received by the candidate.

Specify keys and junction tables

8. After splitting the fields into tables, we must add key fields and junction tables so that we can tie together information in different tables. Otherwise, for example, there would be no way for us to know districts go with which elections.

   For each table, we also must designate a primary key (single field) or composite key (multiple fields taken together) that can be used to uniquely identify each record.

   Newly added fields and tables are highlighted in yellow below.

   a. In Excel, create 3 copies of the original *Senate* sheet. Name each copied sheet after a table from below. Edit each sheet so it only contains the columns (fields) that are appropriate for that table.

Table: *Districts*		
Field Name	**Type**	**Description**
RedistrictingCycle	Number – Integer	Part of composite key. Census data year used for redistricting.
District	Number – Integer	Part of composite key. Number of State Senate district.
Counties	Short Text	List of counties or parts of counties in the district.
Population	Number – Integer	Population of district.

Database Design
WV Senate Problem

Table: *Elections*		
Field Name	**Type**	**Description**
RedistrictingCycle	Number – Integer	Redistricting cycle this election used.
District	Number – Integer	Part of composite key. Number of State Senate district.
Year	Number – Integer	Part of composite key. Year of election.
IncumbentFirstName	Short Text	First name of incumbent.
IncumbentLastName	Short Text	Last name of incumbent.
IncumbentRetired	Yes/No	Indication if the incumbent was not running.
WinnerFirstName	Short Text	First name of the winner.
WinnerLastName	Short Text	Last name of the winner.
WinnerPctg	Number – Percentage	Percentage of the total vote received by the winner.

Table: *Candidates*		
Field Name	**Type**	**Description**
District	Number – Integer	Part of composite key. Number of State Senate district.
Year	Number – Integer	Part of composite key. Year of election.
AmericanFreedomCandidateFirstName	Short Text	First name of the candidate.
AmericanFreedomCandidateLastName	Short Text	Last name of the candidate.
AmericanFreedomRaised	Currency	Amount of funds raised by the candidate.
AmericanFreedomVotes	Number – Integer	Votes received by the candidate.
ConstitutionCandidateFirstName	Short Text	First name of the candidate.
ConstitutionCandidateLastName	Short Text	Last name of the candidate.
ConstitutionRaised	Currency	Amount of funds raised by the candidate.
ConstitutionVotes	Number – Integer	Votes received by the candidate.
DemocraticCandidateFirstName	Short Text	First name of the candidate.
DemocraticCandidateLastName	Short Text	Last name of the candidate.
DemocraticRaised	Currency	Amount of funds raised by the candidate.
DemocraticVotes	Number – Integer	Votes received by the candidate.
LibertarianCandidateFirstName	Short Text	First name of the candidate.

Database Design
WV Senate Problem

Table: *Candidates*		
Field Name	**Type**	**Description**
LibertarianCandidateLastName	Short Text	Last name of the candidate.
LibertarianRaised	Currency	Amount of funds raised by the candidate.
LibertarianVotes	Number – Integer	Votes received by the candidate.
MountainCandidateFirstName	Short Text	First name of the candidate.
MountainCandidateLastName	Short Text	Last name of the candidate.
MountainRaised	Currency	Amount of funds raised by the candidate.
MountainVotes	Number – Integer	Votes received by the candidate.
NaturalLawCandidateFirstName	Short Text	First name of the candidate.
NaturalLawCandidateLastName	Short Text	Last name of the candidate.
NaturalLawRaised	Currency	Amount of funds raised by the candidate.
NaturalLawVotes	Number – Integer	Votes received by the candidate.
RepublicanCandidateFirstName	Short Text	First name of the candidate.
RepublicanCandidateLastName	Short Text	Last name of the candidate.
RepublicanRaised	Currency	Amount of funds raised by the candidate.
RepublicanVotes	Number – Integer	Votes received by the candidate.
WriteInCandidateFirstName	Short Text	First name of the candidate.
WriteInCandidateLastName	Short Text	Last name of the candidate.
WriteInRaised	Currency	Amount of funds raised by the candidate.
WriteInVotes	Number – Integer	Votes received by the candidate.

Database Design
WV Senate Problem

Address redundant or inflexible fields

9. Look for cases where multiple fields store similar information. See if the tables can be redesigned to provide more flexibility.

    Currently, the *Candidates* table contains political party specific fields. The result is that numerous fields are empty and can cause confusion. By reorganizing the table, we can provide more flexibility. While, in most cases, the combination of district, year, and party can uniquely identify each record, it is possible for there to be multiple write-in candidates. For this situation, we must also include the candidate first name and last name fields in the composite key.

    Changes to the table are highlighted in yellow below.

    a. In Excel, edit the data on the *Candidates* sheet to that it matches the format shown below.

Table: *Candidates*		
FieldName	**Type**	**Description**
District	Number – Integer	Part of composite key. Number of State Senate district.
Year	Number – Integer	Part of composite key. Year of election.
PartyName	Short Text	Part of composite key. Political party of the candidate.
CandidateFirst	Short Text	Part of composite key. First name of the candidate.
CandidateLast	Short Text	Part of composite key. Last name of the candidate.
Raised	Currency	Amount of funds raised by the candidate.
VotesReceived	Number – Integer	Votes received by the candidate.

10. Now that we have consolidated our fields, we want to look for cases where we have complicated composite keys. In situations where you must use the combination of multiple fields to uniquely identify a record, it can sometimes be easier to have a single AutoNumber-type field serve as a primary key instead. For example, if we wanted to uniquely identify a person, would it be easier to use their Social Security number or the combination of their name, date and time of birth, location, parents, gender, weight, and height to be sure we are looking at the correct person?

Database Design
WV Senate Problem

a. In the *Candidates* tables, there are 4 fields that collectively serve as a composite key to uniquely identify candidates. It will be easier to instead have a single AutoNumber-type **CandidateID** field serve as primary key.

In Excel, edit the data on the *Candidates* sheet by inserting a new **CandidateID** column and sequentially numbering each candidate (e.g., 1, 2, 3).

Table: *Candidates*		
FieldName	**Type**	**Description**
CandidateID	Number – AutoNumber	Primary key. Unique identifier for the candidate.
District	Number – Integer	Number of State Senate district.
Year	Number – Integer	Year of election.
PartyName	Short Text	Political party of the candidate.
CandidateFirst	Short Text	First name of the candidate.
CandidateLast	Short Text	Last name of the candidate.
Raised	Currency	Amount of funds raised by the candidate.
VotesReceived	Number – Integer	Votes received by the candidate.

11. Identify cases where similar information is repeatedly entered. Fields where users repeatedly enter large amounts of text can be prone to errors such as misspellings.

In this example, the **PartyName** field in the *Candidates* table is a likely source of errors since party names are spelled out. A better solution is to store the party abbreviation as a 3-letter code in the *Elections* table, and then create a *Parties* table to store party abbreviations and their full names.

Changed fields are highlighted in yellow below. The *Parties* table is entirely new.

a. In Excel, create a new sheet to represent the *Parties* table.

Table: *Parties*		
Field Name	**Type**	**Description**
PartyAbbrv	Short Text	Primary key. Abbreviation of the party name.
PartyName	Short Text	Full text of the political party name.

b. Enter records for the parties below.

> **HINT:** The *Parties* table will contain 8 records.

PartyAbbrv	PartyName
AFP	American Freedom Party
CON	Constitution Party
DEM	Democratic Party
LIB	Libertarian Party
MTN	Mountain Party
NLP	Natural Law Party
GOP	Republican Party
WRN	Write-In Candidate

c. Update the *Candidates* table as shown below.

Table: *Candidates*		
Field Name	**Type**	**Description**
CandidateID	Number – AutoNumber	Primary key. Unique identifier for the candidate.
District	Number – Integer	Number of State Senate district.
Year	Number – Integer	Year of election.
PartyAbbrv	Short Text	Political party abbreviation.
CandidateFirst	Short Text	First name of candidate.
CandidateLast	Short Text	Last name of candidate.
Raised	Currency	Amount of funds raised by candidate.
VotesReceived	Number – Integer	Votes received by candidate.

Remove duplicate records

12. After your data has been split into multiple tables, look at your tables to see if there are any duplicate records such as multiple records in the *Districts* table for the same cycle. These should be eliminated.

    a. In Excel, delete any duplicate records that are identical to another record.

Database Design
WV Senate Problem

Review final table structure and cleanup old data

13. Now that we have finished splitting the data into multiple tables and reorganizing the data to avoid duplication, take a moment to review the final table structure shown below.

   a. In Excel, delete the original *Elections* data sheet as it is no longer needed.

Table: *Districts*		
Field Name	**Type**	**Description**
RedistrictingCycle	Number – Integer	Part of composite key. Census data year used for redistricting.
District	Number – Integer	Part of composite key. Number of State Senate district.
Counties	Short Text	List of counties or parts of counties in the district.
Population	Number – Integer	Population of district.

Table: *Elections*		
Field Name	**Type**	**Description**
RedistrictingCycle	Number – Integer	Redistricting cycle this election used.
District	Number – Integer	Part of composite key. Number of State Senate district.
Year	Number – Integer	Part of composite key. Year of election.
IncumbentFirstName	Short Text	First name of incumbent.
IncumbentLastName	Short Text	Last name of incumbent.
IncumbentRetired	Yes/No	Indication if the incumbent was not running.
WinnerFirstName	Short Text	First name of the winner.
WinnerLastName	Short Text	Last name of the winner.
WinnerPctg	Number – Percentage	Percentage of the total vote received by the winner.

Database Design
WV Senate Problem

Table: *Candidates*		
Field Name	**Type**	**Description**
CandidateID	Number – AutoNumber	Primary key. Unique identifier for the candidate.
District	Number – Integer	Number of State Senate district.
Year	Number – Integer	Year of election.
PartyAbbrv	Short Text	Political party abbreviation.
CandidateFirst	Short Text	First name of candidate.
CandidateLast	Short Text	Last name of candidate.
Raised	Currency	Amount of funds raised by candidate.
VotesReceived	Number – Integer	Votes received by candidate.

Table: *Parties*		
Field Name	**Type**	**Description**
PartyAbbrv	Short Text	Primary key. Abbreviation of party name.
PartyName	Short Text	Full text of political party name.

Determine relationships

14. Now that the tables have been defined, we must determine the exact relationships that will tie the tables together. Relationships between two tables are based on (one or more) common fields that appear in both tables. All of the tables must be interconnected. By following one or more relationships, it should be possible to tie data from one table together with data in any other table.

Table 1	Table 2	Common Fields for Relationship
Parties	*Candidates*	**PartyAbbrv**
Districts	*Elections*	**RedistrictingCycle** and **District**
Elections	*Candidates*	**District** and **Year**

Grading Rubric

This assignment is worth 8 points. It will be graded by your instructor using this rubric:

Standard	Meets Requirements (8 points)	Does Not Meet Requirements (0 points)
Student made reasonable effort in correctly completing assignment.	Assignment is at least 70% complete and correct, or student contacted instructor for help on incorrect or incomplete items.	Assignment is less than 70% complete and correct, and student did not contact instructor for assistance on incorrect or incomplete items.

Database Design
WV Senate Problem

This rubric will be used for peer evaluation of this assignment:

Standard	Excellent	Satisfactory	Needs Improvement
Assignment is correct and complete.	Assignment is at least 90% complete and correct.	Assignment is 70%-89% complete and correct.	Assignment is less than 70% complete and correct.

References

[1] D. E. Holmes, *West Virginia Blue Book 1999*, vol. 81. Charleston, West Virginia: West Virginia Legislature, 1999.

[2] D. E. Holmes, *West Virginia Blue Book 2008*, vol. 90. Charleston, West Virginia: West Virginia Legislature, 2008.

[3] D. E. Holmes, *West Virginia Blue Book 2012*, vol. 90. Charleston, West Virginia: West Virginia Legislature, 2012. Available: *http://www.legis.state.wv.us/legisdocs/2012/bluebook/bluebook2012.pdf*.

[4] "Elections - History & Data," *West Virginia Secretary of State*. Available: *http://www.sos.wv.gov/elections/history/Pages/default.aspx*.

[5] "Campaign Finance Reporting System," *West Virginia Secretary of State*. Available: *http://cfrs.wvsos.com/#/home*.