Spreadsheets: Techniques for Managing and Checking Data
How to Use this Guide

This handbook accompanies the taught sessions for the course. Each section contains a brief overview of a topic for your reference and then one or more activities.

If you have any problems with the text or the exercises, please ask the teacher or one of the demonstrators for help.

This book includes plenty of activities and pointers to other sources of information – more than can usually be completed during the hands-on sessions of the course. The teacher and demonstrator(s) are around to guide you and you are encouraged to ask questions and seek advice. Later, you may attend follow-up sessions at ITLP called Computer8, where you can continue work on the activities, with support from IT teachers. Other options include custom courses for teams of people or hourly paid consultancy from the IT learning programme.

Text Conventions

A number of conventions are used to help you to be clear about what you need to do in each step of a task.

- Multiple key names linked by a + (for example, CTRL+Z) indicate that the first key should be held down while the remaining keys are pressed; all keys can then be released together.

- Drop-down menu options are indicated by the name of the options separated by a vertical bar, for example File|Print. In this example you need to select the option Print from the File menu or tab. To do this, click when the mouse pointer is on the File menu or tab name; move the pointer to Print; when Print is highlighted, click the mouse button again.

- A button to be clicked will look like this.

- The names of software packages are identified like this, and the names of files to be used like this.
Software Used

Excel 2010

Files Used

Revision Information

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Author</th>
<th>Changes made</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>Jan 2014</td>
<td>Steven Albury</td>
<td>Initial version</td>
</tr>
<tr>
<td>1.1</td>
<td>Oct 2015</td>
<td>Steven Albury</td>
<td>updates</td>
</tr>
</tbody>
</table>

Copyright

Steven Albury makes this document and the accompanying material available under a Creative Commons licence: Attribution, Non-Commercial, No Derivatives. Individual resources are subject to their own licencing conditions as listed.

Screenshots in this document are copyright of Microsoft.

The Oxford University logo and crest is copyright of Oxford University and may only be used by Oxford University members in accordance with the University’s branding guidelines.
Contents

1 Introduction .................................................................................................................. 5
  1.1. What You Should Already Know ........................................................................ 5
  1.2. What Will You Learn? ......................................................................................... 5
  1.3. What is Excel? ......................................................................................................... 6
  1.4. Where Can I Get A Copy? .................................................................................. 6
  1.5. Other Tools .............................................................................................................. 6
2 Organising your data ..................................................................................................... 7
  2.1. 5 Good Reasons to Organise your data with tables ............................................. 7
  2.2. Creating a table ....................................................................................................... 7
  2.3. Cleaning up your data ............................................................................................. 9
    2.3.1. Widely Used Techniques for cleaning your data ............................................. 9
    2.3.2. The ‘copy column’ technique ....................................................................... 9
  2.4. The data cleaning process: .................................................................................... 9
3 Adapting and Managing Existing Spreadsheets ......................................................... 11
  3.1. Spreadsheet testing process ................................................................................... 11
    3.1.1. Ensuring data has been cleansed and spreadsheet has ‘face validity’ ............ 11
    3.1.2. Testing process step 1 .................................................................................. 12
    3.1.3. Testing process step 2 .................................................................................. 12
    3.1.4. Testing process step 3 .................................................................................. 12
    3.1.5. Testing process step 4 .................................................................................. 12
    3.1.6. Testing process step 5 .................................................................................. 12
4 Planning and Designing New Spreadsheets ............................................................... 13
  4.1. Planning and design ............................................................................................... 13
  4.2. Use Cases .............................................................................................................. 14
  4.3. Absolute and relative cell references .................................................................. 14
  4.4. Testing testing testing ......................................................................................... 14
5 Other Courses .................................................................................................................. 18
  5.1. The ITLP Portfolio ................................................................................................... 18
  5.2. Other Spreadsheet Courses .................................................................................. 18
  5.3. Courses in Statistical Analysis ............................................................................ 18
  5.4. Courses in Programming Languages .................................................................... 18
  5.5. Computer8 ............................................................................................................. 19
  5.6. IT Services Help Centre ....................................................................................... 19
6 Slide Presentation ........................................................................................................... 20
7 Links and References .................................................................................................... 30
1 Introduction

Welcome to the Excel: Managing and Checking Data workbook!

This booklet accompanies the course delivered by Oxford University’s IT Learning Programme. The course includes a number of group activities and is intended for classroom use but you may also decide to review it in your own time. At the back of the book is a copy of the slides used during the session.

If at any time you are not clear about any aspect of the course, please make sure you ask your teacher or demonstrator for some help. If you are away from the class, you can get help by email from your teacher or from help@it.ox.ac.uk.

1.1. What You Should Already Know

No advanced knowledge of Excel is expected but you should be familiar with the Excel interface.

It is also assumed that you are familiar with opening files from particular folders and saving them, perhaps with a different name, back to the same or a different folder. Please ask the sessional teacher if this is something you need help with.

The computer network in our teaching rooms may differ slightly from that which you are used to in your College or Department; if you are confused by the differences please ask for help from the teacher or demonstrator(s).

1.2. What Will You Learn?

This course will help you learn to use data management concepts when planning creating and maintaining Excel spreadsheets. While there are specific exercises relating to Microsoft Excel the concepts of planning and organising data are transferrable to other spreadsheet tools such as ‘Numbers’ and more broadly to other data management applications such as SPSS or NVivo.

The following topics are the main ones covered in the course:

- Organising your data
- Documenting and testing as you go
- Auditing a spreadsheet
- Making changes to a complex workbook
- Designing a solution to your problem

These notes are written with Microsoft Excel for Windows in mind. Having worked through these notes, you should also be able to adapt to other versions including Excel for Mac, since most of the principles hold true regardless of the version of the software.

Getting to grips with a package as sophisticated and powerful as Excel can be time-consuming, so allow yourself plenty of time for practice.

In these notes, topics marked Optional are additional material which may not necessarily be covered by your lecturer in the taught course. Depending on the work you need to use Excel for, you may find such topics helpful. For further information on these, consult Excel’s Help.
1.3. What is Excel?

Spreadsheets allow you to present and analyse data, especially numerical data, in a wide variety of ways. Microsoft Excel is perhaps the world’s most popular spreadsheet application and is a core application in the Microsoft Office suite.

1.4. Where Can I Get A Copy?

If you have a copy of Microsoft Office 2010, then you already have a copy of Excel 2010. If you are unable to find it on your computer, it may not have been installed and you should talk to your IT support contact (or the IT Services Help Centre).

If you are a member of staff, you can obtain a copy of Microsoft Office from the IT Services on-line shop. Students can obtain a Microsoft Student Licence, but this must be bought through a Microsoft Authorised Education Reseller.

1.5. Other Tools

There are many other spreadsheet tools that you could use. For example a popular tool on the Mac is Numbers. Although these tools will differ in the way that the features are used, the principles described in this session are equally applicable.

An Open Source alternative is LibreOffice Calc, available for Windows, Mac and Linux computers.

There are also a number of web based tools that you might investigate. For example, Google Docs spreadsheet

If you use a tool other than Excel and you are unsure how you might create a particular aspect of your presentation, the local IT officer in your college or department is the person to contact in the first instance and then the IT Services Help Centre if necessary.
2 Organising your data

There are many approaches to organising your work. In this section there is a description of some of the things that have been found useful by other people in the university.

In the classroom we will carry out some activities to help you develop your skills in this area. A lot of these will involve using tables.

2.1. 5 Good Reasons to Organise your data with tables

1. Sorting and filtering
2. Easy to add new data – extends table rows and columns easily
3. Formulas adjust to take account of new rows and columns
4. Can always see table header – helps avoid entering data into wrong column
5. Auto-complete formulas – avoid accidentally not copying a formula correctly

2.2. Creating a table

If you have not created a table before it can be done as follows:

On the worksheet you wish to use, select ONE CELL containing data:

<table>
<thead>
<tr>
<th>TE</th>
<th>FIRM CODE</th>
<th>COMPANY NAME</th>
<th>BUDGET CODE</th>
</tr>
</thead>
<tbody>
<tr>
<td>310</td>
<td>1</td>
<td>A.Daley Inc</td>
<td>4</td>
</tr>
<tr>
<td>221</td>
<td>3</td>
<td>Hardcopy stores</td>
<td>4</td>
</tr>
<tr>
<td>359</td>
<td>1</td>
<td>A.Daley Inc</td>
<td>4</td>
</tr>
<tr>
<td>396</td>
<td>5</td>
<td>PerfectWard</td>
<td>4</td>
</tr>
<tr>
<td>144</td>
<td>2</td>
<td>Comp. Stores</td>
<td>4</td>
</tr>
<tr>
<td>214</td>
<td>1</td>
<td>A.Daley Inc</td>
<td>4</td>
</tr>
<tr>
<td>227</td>
<td>5</td>
<td>PerfectWard</td>
<td>4</td>
</tr>
<tr>
<td>303</td>
<td>4</td>
<td>Jim Bowen Ltd</td>
<td>4</td>
</tr>
<tr>
<td>312</td>
<td>5</td>
<td>PerfectWard</td>
<td>4</td>
</tr>
<tr>
<td>153</td>
<td>1</td>
<td>A.Daley Inc</td>
<td>4</td>
</tr>
<tr>
<td>178</td>
<td>3</td>
<td>Hardcopy stores</td>
<td>4</td>
</tr>
<tr>
<td>200</td>
<td>3</td>
<td>Hardcopy stores</td>
<td>4</td>
</tr>
<tr>
<td>124</td>
<td>4</td>
<td>Jim Bowen Ltd</td>
<td>4</td>
</tr>
<tr>
<td>303</td>
<td>4</td>
<td>Jim Bowen Ltd</td>
<td>4</td>
</tr>
</tbody>
</table>
From the ‘Insert’ tab on the ribbon select **Table**

A dialog box will appear suggesting the area to be included

Check the highlighted area matches what you intended and click OK.

A table will now be created and an autoformat applied to it. This can be changed from the ‘design’ tab
2.3. Cleaning up your data

There are certain activities that should be seen as good practice when preparing an Excel worksheet for use. This is particularly important when you have imported the data from an external source. The following guide highlights the main areas where text and numerical data can be checked for errors and put into a state where analysis can be undertaken.

2.3.1. Widely Used Techniques for cleaning your data

- Use spell check – set default for UPPER case – usual caveats apply
- Look for duplicate rows – make sure you only have data once
- Finding and replacing text and formatting
- Fixing numbers and number signs
- Fixing dates and times – especially things such as SPSS dates
- Merging and splitting columns
- Transforming and rearranging columns and rows

2.3.2. The ‘copy column’ technique

When cleaning up data you may want to use Excel functions such as UPPER() to convert into upper case or TRIM() to remove leading spaces from a cell. A good way to do this is to insert a new column next to the one you want to clean up. In that column use the cleaning function e.g. PROPER() to put text into sentence case and then select the cell needing cleaning. At this point you have two columns. Copy the corrected one and then paste it back into the same column BUT paste values rather than the formula. You can then delete the original column with the erroneous text/formatting.

2.4. The data cleaning process:

This is a suggested structured approach to cleaning up your data and should help ensure you get rid of formatting and invalid cell entries. This then needs to be taken a step further by testing the formulas and the validation rules for data entry. The final stage is to then protect the worksheet to avoid accidental or malicious changes to data. After data cleansing you should follow the process for validating and protecting data entry to help avoid future problems. More information on protecting worksheets and on validating data entry can be found in other ITLP courses and you should attend these if you need to build up skills in these areas (see section 5 for further information.)
1. **ALWAYS** Create a backup copy of the original data in a separate workbook.
2. Ensure the data is organised into named rows and columns – a table is often useful.
3. Do tasks (e.g. spell-checking or finding and replacing text) that don’t require copy-column changes first.
4. Move on to making changes that require the copy-column technique. This includes removing spaces, adjusting date formats (especially if imported from an external source such as SPSS).

The general steps for doing this are:

1. Insert a new column (B) next to the original column (A) that needs cleaning.
2. Add a formula that will transform the data at the top of the new column (B).
3. Fill down the formula in the new column (B). In an Excel table, a calculated column is automatically created with the values filled down.
4. Select the new column (B), copy it, and then paste as values into the new column (B).
5. Remove the original column (A), which converts the new column from B to A.

To periodically clean the same data source, consider recording a macro or writing code to automate the entire process. There are also a number of external add-ins written by third-party vendors, listed in the Third-party providers section, that you can consider using if you don’t have the time or resources to automate the process on your own.
3 Adapting and Managing Existing Spreadsheets

One of the most common situations where errors occur is when a spreadsheet is created and then used over a much longer time period than originally envisaged. This can be compounded by the original creator moving on to a different role and the inherited spreadsheet being relied on by users who had no part in its original creation.

In this section there is an outline of a validation and testing process which might be helpful in identifying and correcting errors as well as showing how spreadsheets can be simplified and changes to them can be managed and logged.

Also in this section there is a description of how spreadsheets can be protected from accidental change and data entry can be restricted to pre-defined ranges.

3.1. Spreadsheet testing process

The following explains a process for testing spreadsheets and also provides examples of key areas to test and techniques to use. Testing can be both a time consuming and mundane activity. The importance of having an ongoing and formalised testing process, where results are recorded within the spreadsheet cannot be overstated. However at the same time a balance should be struck between writing test scripts, which can take a lot of time and carrying out testing itself. The following solution is based on one proposed for use within the pharmaceutical industry and can be adapted to a wide range of applications.

The important aspect of testing that is often overlooked is that it is intended to act as a snapshot at a single period of time and to confirm the status of formulas and other cell entries at that time only. This means that a post-implementation audit process is also needed as well as testing while the spreadsheet is being developed.

3.1.1. Ensuring data has been cleansed and spreadsheet has ‘face validity’

This first step in the process is a common sense check. There should be a record within the spreadsheet on a worksheet named ‘documentation’ of when the data was checked for correct data types (e.g. dates) and the other data cleansing operations described above. It is also a good idea to ensure a named person has responsibility for this and that they have signed it off. If you are a researcher working on your own then clearly a date when you completed tidying up the workbook will be sufficient.

In addition to this before testing formally begins you should have a quick check to ensure that it looks as though all the necessary data is there or where gaps are present an agreed placeholder such as ‘null’ or ‘unavailable’ has been used. This check should also reveal whether any obvious errors are present such as a DIV/0! Error.
If any such errors are present they should be corrected prior to the formal testing process beginning.

**Using IF() and ISERROR**

It is sometimes quite sensible for there to be a zero (0) in a cell which is in a column where division is being carried out. The expected result of such a calculation will be zero. However Excel will produce an error #DIV/0!. To avoid this you can create a default entry for cells in any column where this might happen. A trivial example is where the cost of something is being divided amongst users. For things where there are no users but there are costs it will be necessary to plan for coping with DIV/0 errors.

By combining the functions IF() and ISERROR you can set such cells to 0 or any other default value.

The formula to use can be based on this:

IF(ISERROR(A1/B1), 0, A1)

This will then become a formula which is covered in the testing process itself.

**3.1.2. Testing process step 1**

Test the structure. This means going through formulas and checking that they are working on the correct values.

**3.1.3. Testing process step 2**

Test the calculations. Having ensured the correct values are being used you can now test that the correct formula (and associated functions) is correct.

**3.1.4. Testing process step 3**

Test any macros/automation. This needs to be done in a highly methodical manner as macros can become complex quite quickly and can involve multiple pathways as choices are made.

**3.1.5. Testing process step 4**

Having ensured the spreadsheet is correct in terms of the formulas present, the data values in cells and any automation you can test each use case. It should be possible to go through each use case and complete it.

**3.1.6. Testing process step 5**

The spreadsheet has now been tested in terms of both requirements (it does what is required of it) and function (the formulas have been tested, data entry validated and errors such as DIV/0! handled properly. A final step is to ensure that calculations can be repeated manually. The manual testing provides a protection against drag and drop errors where a mistake in formula construction has been missed. By calculating the results of formulas manually using paper and a calculator or other technique outside of the spreadsheet it can be guaranteed that the results are accurate.
4 Planning and Designing New Spreadsheets

Planning and designing new spreadsheets can be done at a number of levels of detail. It is not good practice to have a single process as the things that a spreadsheet might be used for are so varied. Instead a risk-based approach should be adopted with the level of formal process being determined by the risk weighting attached to spreadsheet errors.

Calculating such risks is dependent on the environment and the potential for various types of risk (e.g. reputational, financial, safety and technical) to occur and the potential likelihood of any one of them occurring.

The following process therefore should be seen as a description of a more formal but flexible approach and should be adapted to accord with the needs of all the spreadsheets stakeholders (a group which might include study participants who are entitled to the protection of personal data; the wider university where a serious error can lead to embarrassment and potentially to loss of income).

A short case study in the problems caused by not adopting a risk based approach is provided by one of the faculties here at the University.

An unnamed faculty within the collegiate university which supports the study of events that happened in the past and teaches undergraduates how to analyse situations that occurred in the past developed a spreadsheet to manage an entrance exam. The spreadsheet was clearly labelled and was only using straightforward formulae. However, there was no validation or protection of worksheets. Each tutor completed their own version of the sheet and these were then consolidated into a master sheet. Some tutors transposed the student id number and the score. These data were imported into the master sheet and it only later became clear there was a problem. However, the embarrassment and reputational damage to the faculty was substantial. Despite the importance of the output from that Excel workbook it was treated as ‘just a spreadsheet’.

The importance of treating spreadsheets as a core support tool requiring planning and support is an important element of a strategy to reduce errors and increase the confidence that can be placed in reported results.

4.1. Planning and design

The following are seven recommendations for use when designing new spreadsheets. They all relate to trying to prevent future errors by making the structure and purpose of all aspects of a workbook clear and easy to follow.

1. Restrict input to valid data using validation options
2. Split out data to an atomic level (use FirstName LastName rather than just Name for example)
3. Use ‘show formulas’ in the formulas|show formulas menu on the ribbon or pressing the Ctrl + ¬` button on the top left of the keyboard – this can help ensure you are putting the right formula in the right location.

4. Click into the formula bar which highlights the cells being used in a formula – this is a type of testing activity that can be carried out as you go along.

5. Use the worksheet protection facility to lock all cells except input cells – this helps prevent errors such as accidental deletion of a formulas which is replaced by a fixed value.

6. Add ‘How To’ and ‘Documentation’ tabs to all workbooks. This provides future users with a place to find out what formulas have been used and how they work as well as a log of changes and a short record of design decisions. This is not easy to keep up and therefore should be a mandated requirement so that it becomes a habit.

7. Use named ranges wherever possible rather than cell references to aid clarity.

### 4.2. Use Cases

A ‘use case’ is a detailed written description of the possible pathways for a user through the spreadsheet. A use case has a name (e.g. ‘Add experiment results’) and a detailed written description of who does the activity and what steps they take to complete it. Use cases should be written to cover all the ways in which the sheet might be used. These can be kept in the ‘documentation’ worksheet.

### 4.3. Absolute and relative cell references

One aspect of Excel which can cause subtle errors is a misuse of absolute and relative cell references. Be sure that when creating a new spreadsheet there is correct use of the sign.

<table>
<thead>
<tr>
<th>Cell reference</th>
<th>Type and effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>relative reference</td>
</tr>
<tr>
<td>$A$1</td>
<td>absolute reference (same as a named cell)</td>
</tr>
<tr>
<td>$A$1</td>
<td>column is absolute and won’t change</td>
</tr>
<tr>
<td>A$1</td>
<td>row is absolute and won’t change</td>
</tr>
</tbody>
</table>

It is recommended that any use of absolute references especially non-standard ones where only the column or row is fixed is fully documented within the workbook.

### 4.4. Testing testing testing

In section three there is a description of a suggested test/audit process. This also applies to new spreadsheets. Prior to creating the sheet itself the requirements should be clearly defined. When the first version of a new sheet is created a test cycle should be included. This is sometimes referred to as test-driven development. There is likely to be a series of build/test/release cycles to any new spreadsheet. The testing part is just as important as the development part.
Spreadsheets: Techniques for Managing and Checking Data
5 Other Courses

Courses offering skills development in spreadsheets and related topics are described below. In all cases, please refer to the IT Learning Programme catalogue (via www.it.ox.ac.uk/courses/) for further details.

5.1. The ITLP Portfolio

These course materials are available through the ITLP Portfolio, at http://portfolio.it.ox.ac.uk.

Each course pack includes the course handbook in pdf form and a zip folder of the exercise files that you need to complete the exercises. Archive versions of the course book may also be useful if you use an earlier version of the software.

The ITLP Portfolio helps you find articles, videos, resources and weblinks for further IT study. For some resources, you will be asked for your Oxford (SSO) username and password.

5.2. Other Spreadsheet Courses

- Excel: Functions and cell referencing
- Excel: 3D formulae, charts and sharing data
- Excel: Analysing spreadsheet external data and what-if analysis

5.3. Courses in Statistical Analysis

- SPSS: An introduction
- R: An introduction
- NVivo: Introductory workshop
- Stata: An introduction to data access and management

Some prerequisite knowledge is required. See the IT Learning Programme catalogue for details http://courses.it.ox.ac.uk/).

5.4. Courses in Programming Languages

For those who need to process large amounts of data, knowledge of a programming language may be useful. IT has large libraries of programs which perform all sorts of numerical analyses and plots. These are easily accessed with a simple program. The following courses are available (See the IT Learning Programme catalogue for details http://courses.it.ox.ac.uk):

- Programming: Concepts
- Programming: Java Introduction
- Programming: C++
- Programming: Perl Introduction
- Programming: Python Introduction
- Introduction to R
5.5. Course Clinic

We encourage everyone to work at their own pace. This may mean that you don’t manage to finish all of the exercises for this session. If this is the case, and you would like to complete the exercises while someone is on hand to help you, come along to one of the course clinic sessions that run during term time.

More details are available from the IT Learning Programme course catalogue (via www.it.ox.ac.uk/courses/).

5.6. IT Services Help Centre

In the IT Services Help Centre, you can use the facilities to work through the exercises in this booklet, or use any of the applications that are available. The Help Centre is also a good place to get advice about any aspect of using computer software or hardware.

For Help Centre opening times, visit www.it.ox.ac.uk/help/gettinghelp/ and follow links to the General Helpdesk, or contact them by email on help@it.ox.ac.uk.
6 Slide Presentation

The following presentation is used as a basis for discussion in the taught sessions for this course – it highlights the key issues to be considered when managing and designing spreadsheets.

Slide 1

Excel: Managing & Checking Data

Slide 2

Today's arrangements

| Your teacher is: Steven Albury |
| We finish by: 12:10pm (or 4:50pm for afternoon course) |
| Please ask questions, raise topics, make a point or share your own experiences - it makes the course more useful and more enjoyable |

Slide 3

Your safety is important

| Where is the fire exit? |
| Beware of hazards |
| Please tell us if anything does not work |
| Let us know if you have any other concerns |
Slide 4

Your comfort is important

- The toilets are along the corridor just outside the teaching rooms
- The rest area is where you registered; it has vending machines and a water cooler
- The seats at the computers are adjustable
- You can adjust the monitors for height, tilt and brightness

Slide 5

Session plan - 3 main topics

- Organising data
- Managing spreadsheets
- Designing new spreadsheets

Slide 6

Excel Problems - Which are the most common?

Think of one problem you have had with Excel that you would like to ensure doesn’t happen again & discuss it with your neighbour
### Slide 7: Organising data

- What are the potential problems with Excel?
- Data cleansing
- Data types and cell referencing
- Structure and clarity
- Validation and overcoming data entry errors

### Slide 8: Organising Data - Background

- Errors in most spreadsheets - 94% of spreadsheets have been found to have errors with 5% of all cells having errors (pwc consulting)
- Need for better traceability under service standards such as ITIL or Research standards e.g. RRI - part of carrying out work consistent with Horizon 2020 standards
- Can undermine research / project credibility

### Slide 9: 6 main types of errors

Powell et al settled for six error types:
- Hard-coding in a formula - one or more "magic" numbers appear in formulas
- Reference error - a formula contains one or more incorrect references to other cells
- Logic error - a formula is used incorrectly, leading to an incorrect result
- Copy/Paste error - a formula is wrong due to inaccurate use of copy/paste
- Omission error - a formula is wrong because one or more of its input cells is blank
- Data input error - an incorrect input is used
Organising your data – Data Cleansing

As well as the 6 common errors there is the problem of data being imported or entered incorrectly. Data cleansing is the process of tidying things up before analysis takes place. However there is always scope for a slow degradation of data quality - there is a need for ongoing testing and maintenance.

Data Cleansing - Activity

In pairs I would like you to identify and fix the data errors in the file “Cars data cleansing.xls” spreadsheet on the H: Drive. You should be able to discuss what process you went through and which Excel tools you used to fix the problem (hint look in the workbook for some clues).

Organising Data: Using Tables

Tables are a way of using Excel in a similar way to a database:

- Covered in detail on the course Excel: Analysing spreadsheet lists external data and what-if-analysis
- Offer a number of significant benefits when organising data
Slide 13

Why use tables?

- Sorting and filtering - automatically added
- Easy to add new data - auto extends table
- Formulas adjust to take account of new rows and columns
- Can always see table header - helps avoid entering data into wrong column
- Auto-complete formulas - avoid accidentally not copying a formula correctly

Slide 14

Questions and ideas

Take aways from this part of the session:

Cleaning up data can save you tonnes of effort later
Following a structured approach means you won’t miss a step
Always worth thinking about the level and extent of organisation needed - BUT don’t underestimate
Slide 16

Managing Spreadsheets

Spreadsheets are often inherited and can be difficult to unravel - this leads to trust being placed in them that may not be warranted. Spreadsheets should therefore be tested, especially in terms of the accuracy of formulas etc.

Slide 17

Auditing an existing spreadsheet

• Protection of sheets and who edited what
• Manual Testing
• Documentation and change management
• Ongoing process - document within the workbook and keep information in one place

Slide 18

Making changes to a complex workbook

• Includes macros and VBA changes - needs a formal change management process (no need to make it heavy duty but needs to be written down and tested) - black box testing means you need to know inputs and desired outputs
• Use cases to model all paths
• Unravelling complex functions
• Checking table lookups
Spreadsheets: Techniques for Managing and Checking Data

Slide 19

Planning and Design

Slide 20

Designing new spreadsheets

A clear idea of the purpose of the spreadsheet and its intended users can save a lot of time later.

Adopt a lightweight design process - don’t ignore the idea of planning and design because it seems too long-winded and tedious:

Slide 21

Key techniques to use when creating new sheets

- Validation - restrict entries to valid data
- Break information down to atomic level (e.g. first name, last name)
- As the design is put together have one person responsible for checking formulas (Ctrl + ` turns spreadsheet into formula mode) as they are developed - this should include manual testing of outcomes
Key techniques to use when creating new sheets

- Use protection of worksheet function - only data entry cells should be editable - BUT document it alongside the design
- Use conditional formatting which can highlight when a value is unexpectedly high/low
- Document the workbook and provide a ‘How To’ guide in named worksheets
- Use tables and named ranges where absolute references are required

Documentation

- Create dependency graphs and check them with the ‘display relationships between formulas and cells’ when testing them
- On documentation worksheet explain workings of all non-trivial formulas – prefer more formulas rather to complex ones

Automation

- A lot of errors occur when data entry or manipulation becomes repetitive so:
  - Use macros where things are done repeatedly
  - Consider writing VBA programs (or hiring a developer for more complex work)
Slide 25

Common Errors in when creating a sheet

- Absolute and relative referencing mixed up
- Accidental replacement of formula with a value (protect worksheets)
- Pointing to wrong cell in a formula so calculation incorrect (really bad if that is then used as basis for another formula!) - named ranges help here

Slide 26

Other courses

Advanced Excel courses:
- Excel: Analysing Spreadsheet Lists - now called Tables
- Excel: Analysing your data with Pivot Tables
- Excel: 3D formulae, charts and sharing data
- Excel: Arrays, macros and VBA

Slide 27

Looking for help?

ITLP Closed courses - tailored courses delivered for specific teams or consultancy at hourly charge

Or for quick questions you can DROP-IN every Friday in term time

Computer8

Come along and get help on specific problems between 10:30 and 12
book online or drop-in
Get All the exercises and lots of other resources - Free to University Card holders

Download the files (and more) from the ITLP Portfolio at http://portfolio.it.ox.ac.uk
Or email help@it.ox.ac.uk

This presentation is made available by Steven Albury under a Creative Commons licence: Attribution-NonCommercial-NoDerivs CC BY-NC-ND steven.albury@it.ox.ac.uk
7 Links and References

Here is a list of some sites that can help you solve Excel problems

http://chandoo.org
www.xlninja.com
http://spreadsheets.about.com/
http://www.youtube.com/user/ExcelIsFun

Any particular technique or trick in Excel can usually be found either at one of these locations or via them.