Data processing in life sciences

Microsoft Excel 2016

Lecture 2

Ferenc Peták

Department of Medical Physics and Informatics

2018

Lecture bonus with Mentimeter

New bonus system

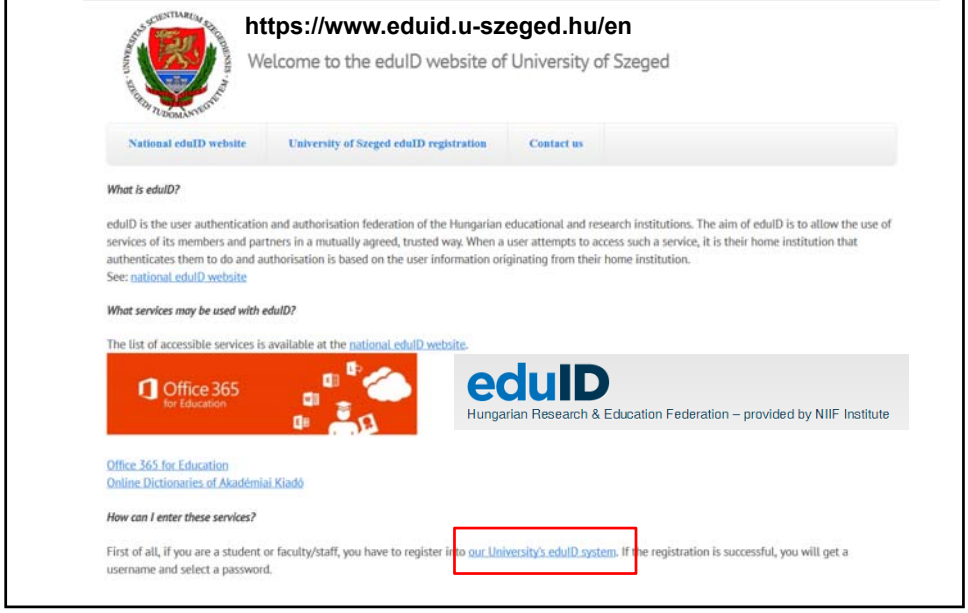
Two bonus points (1%) can be awarded in each lecture if:

1. Providing a valid **full name** at the Mentimeter test
2. >**3 correct answers** out of the 4 online test questions

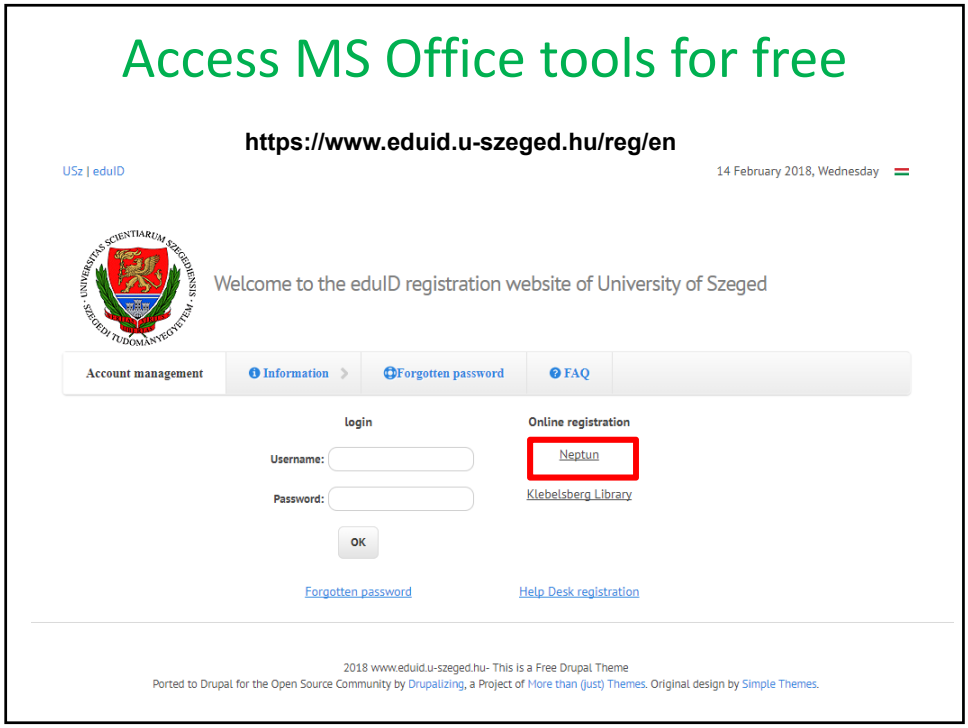
Participation:

- Mobile devices with installed Mentimeter application
- Online device with internet access http://www.menti.com

Access MS Office tools for free

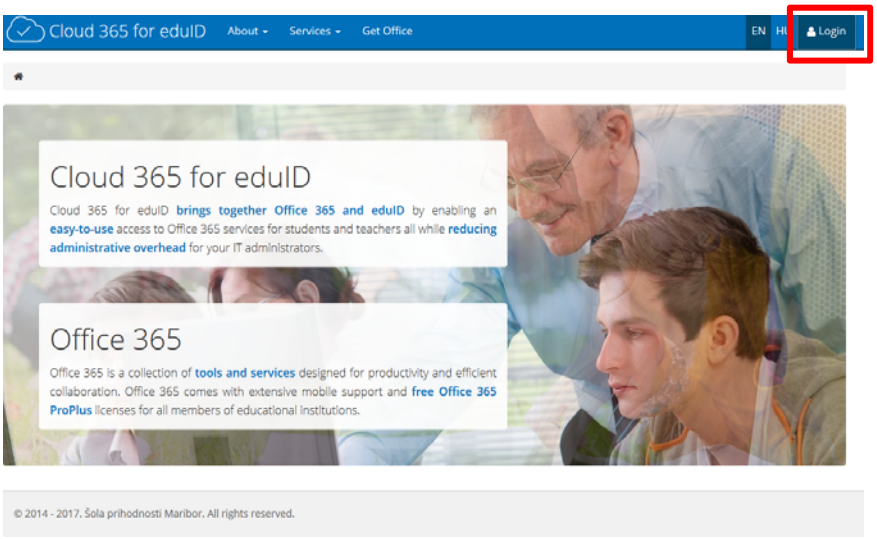


Access MS Office tools for free

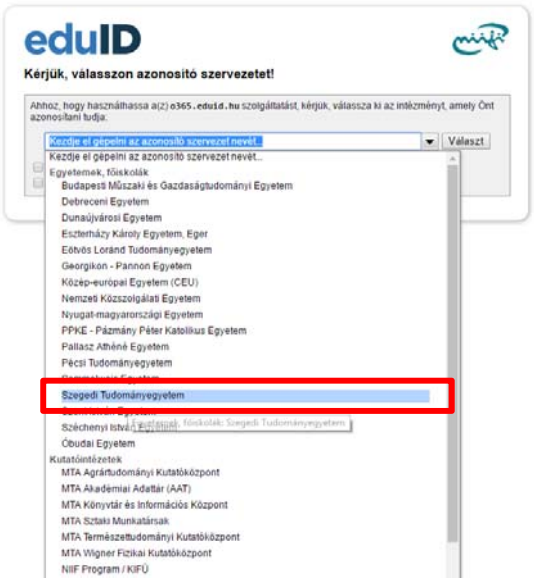


Access MS Office tools for free

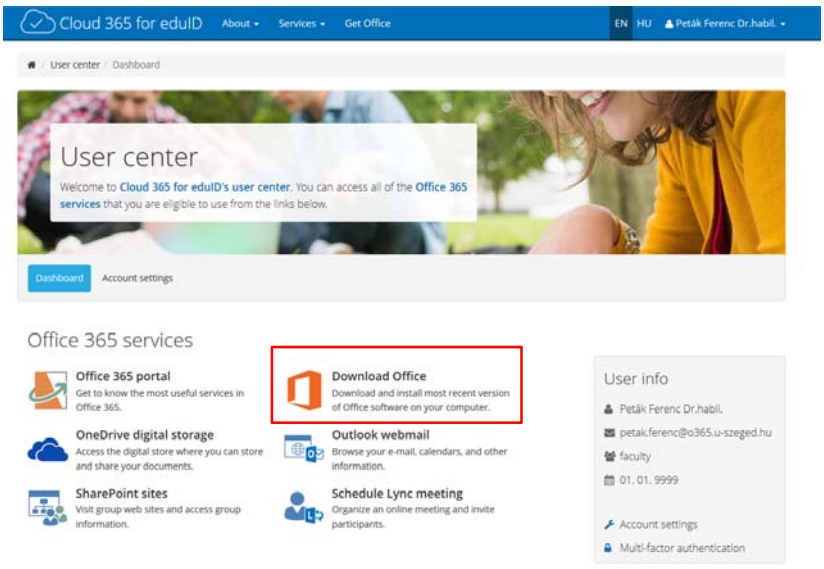
https://o365.eduid.hu/



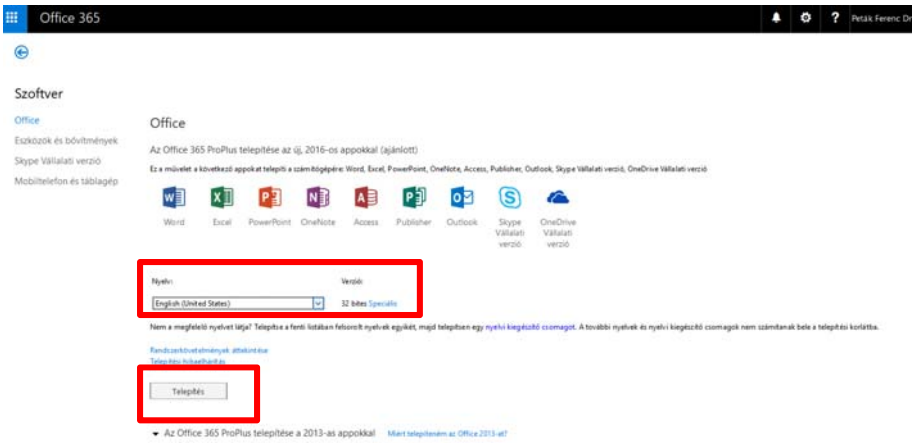
Access MS Office tools for free



Access MS Office tools for free



Access MS Office tools for free



Main topics

1. Formula, reference types
2. Fill handle
3. Excel functions
4. Sorting and filtering
5. Charts
6. Pivot table, pivot chart

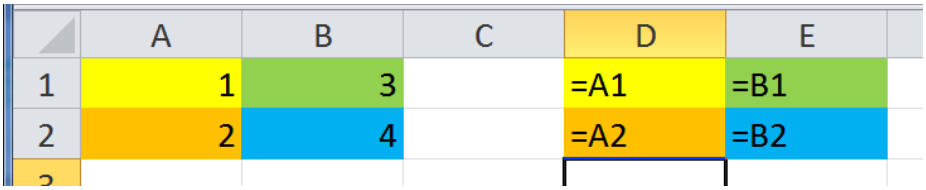
Entering formula – arithmetic operations

- The formula is a mathematical expression that calculates a value
- Formulas always begin with „**=**„ sign
- The order of precedence is a set of predefined rules – in accordance with elementary mathematics
- Operator symbols:
  - Arithmetic: **+ - * / ^ ( )**
  - Comparison: **= > >= < <= <>**
  - Text concatenate: **&**

"Those who can, do. Those who cannot, teach. Those who cannot teach, develop spreadsheets."

Reference types

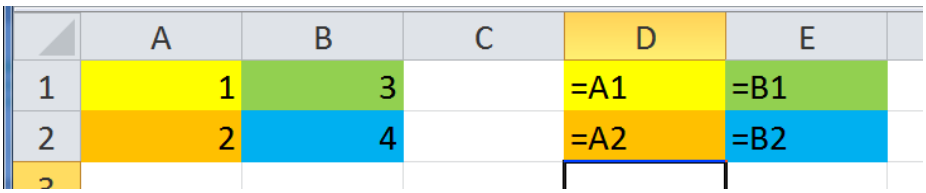
- **A reference identifies a cell or a range of cells on a worksheet**
  - **C13** - the cell in column C and row 13
  - **C1:D10** - the range of cells C1 through D10
  - **C1,D10** – C1 and D10 cells only!
  - **3:3** - all cells in row 3
  - **3:13** - all cells in row 3 through 13
  - **C:C** - all cells in column C
  - **Sheet2!C13** - refers to the worksheet named Sheet2
- References are **not case-sensitive**



	A	B	C	D	E
1	1	3		=A1	=B1
2	2	4		=A2	=B2

Reference types

- **Relative reference: e.g. A1, C3:D6**
  - Based on the relative position of the cell that contains the formula and the cell the reference refers to
  - **Copy → the reference automatically adjusts**

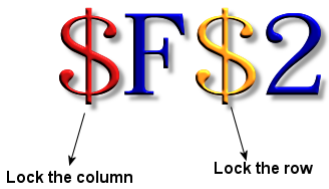


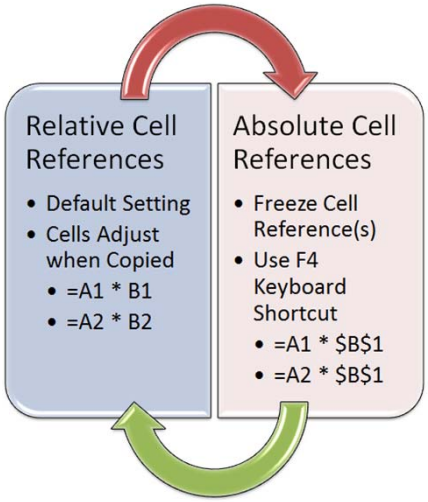
	A	B	C	D	E
1	1	3		=A1	=B1
2	2	4		=A2	=B2

Reference types

- **Absolute reference: e.g. $A$1, $C$3:$D$6**
  - Always refer to a cell in a specific location
  - **The absolute reference does not adjust**
- **Mixed reference: $A1, A$1**
  - Either an absolute column and relative row, or absolute row and relative column If you copy or fill the formula across rows or down columns, **the relative reference automatically adjusts**, and **the absolute reference does not adjust**
- Each push of **F4** will toggle the reference type

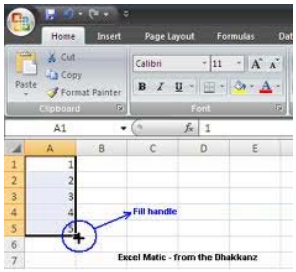
Absolute and relative references



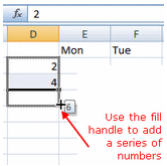


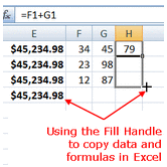
http://www.thecompanyrocks.com/

The fill handle



- **Small black dot or square in the bottom right corner of the active cell**
- The use of the fill handle depends on the active cell content:
  - If the active cell content is a number or text
    - Auto Fill other selected cells with the next items in the series
  - If the active cell content is a formula or function
    - Copy cell content to adjacent cells
- Use: click on it with the mouse pointer to drag select adjacent cells.
- Special use:
  - Double click: fill until the last cell on the left





Use Excel functions

- The Excel function is a predefined (built-in) formula for commonly used data processing
- General syntax of the Excel function:

=FunctDef(argument$_1$, argument$_2$ … argument$_n$)

Defines the function has to be performed
e.g. ***=average***
=count
=stdev

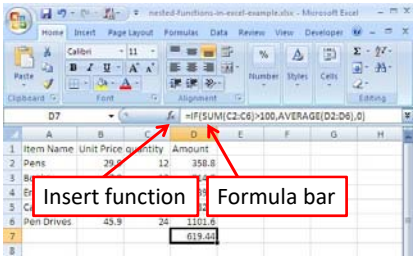
Numbers, text, cell references used by the function
Some arguments are optional
Some functions has no arguments

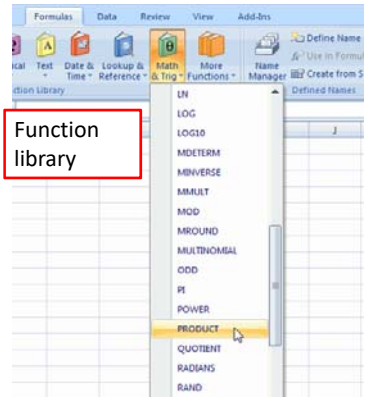
„There are two ways to develop error-free spreadsheets. Only the third one works. „

Use Excel functions

- Excel supplies more than 350 functions organized into 10 categories
  - Database, Date and Time, Engineering, Financial, Information, Logical, Lookup, Math, Text and Data and Statistical
- Entering a function:
  - Manual (typing) – risk for a mistake
  - Insert function button on the Formula bar
  - Furmulas tab – Function library group

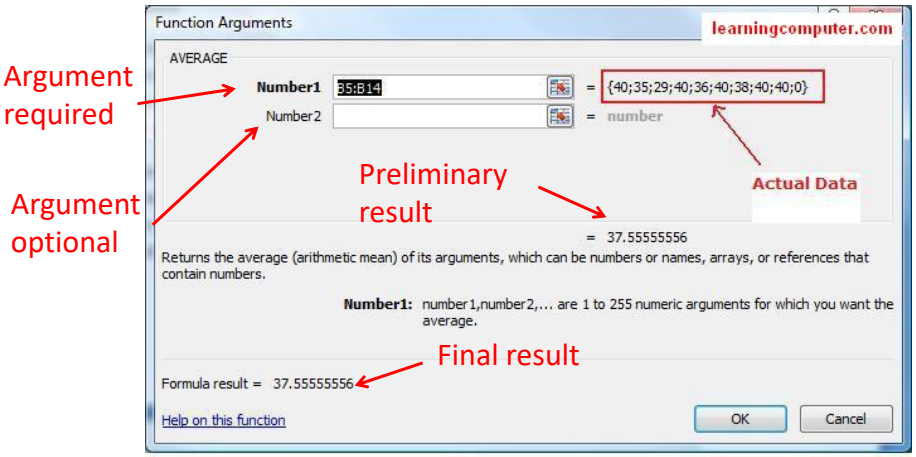
Recommended to avoid mistakes and have description





Use Excel functions

After selecting the appropriate function



Sorting, ranking

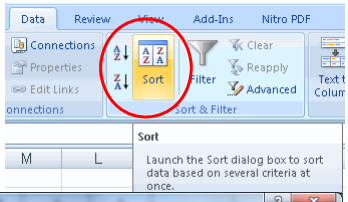
Why sorting?

- Easier to follow the sorted data
- Easier to find a specific information (name, group, treatment, outlier data, etc)
- Helps discovering data entry errors
- Excel help sorting in ascending (increasing) or descending (decreasing) orders

Sorting

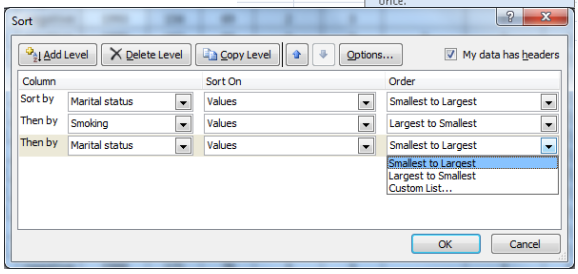
Before sorting: make sure: all data are selected, or a cell is selected within a continuous table!

Tip: a continuous table is sorted based on the category of the selected cell



Made in 2 steps:

1. Select data in the worksheet to be sorted (sorting will be based on the current column)
2. Data/Sort and Filter
   - Ascending
   - Descending order



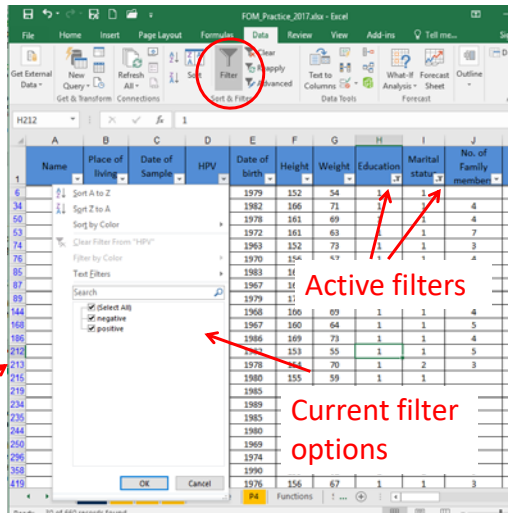
Filtering

- Working with a subset of data selected form the original data base
- A filtered range displays only the rows that meet the criteria
- Unlike sorting, filtering does not rearrange the data
- Temporarily hides the rows that are not intended to display

- Select the Filter option, from the Sort & Filter ribbon in the Data tab.

Filtering - Autofilter

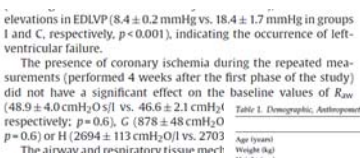
- Find and work with a subset of data
- After filtering
  - reapply a filter to get up-to-date results
  - clear a filter to redisplay all of the data.



Working with charts

Quantitative data in life sciences can be represented as:

Text


elevations in EDLVP (8.4 ± 0.2 mmHg vs. 18.4 ± 1.7 mmHg in groups I and C, respectively, $p < 0.001$), indicating the occurrence of left-ventricular failure.

The presence of coronary ischemia during the repeated measurements (performed 4 weeks after the first phase of the study) did not have a significant effect on the baseline values of R_{aw} (48.9 ± 4.0 cmH_2O s/l vs. 46.6 ± 2.1 cmH_2 respectively; $p = 0.6$), G (878 ± 48 cmH_2O $p = 0.6$) or H (2694 ± 113 cmH_2O/l vs. 2703

The airway and respiratory tissue mech

Table

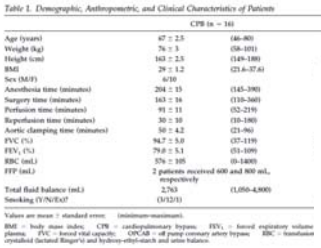


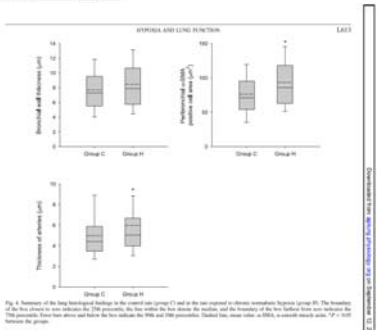
Table 1. Demographic, Anthropometric, and Clinical Characteristics of Patients

	CPB (n = 16)		OPCAB (n = 19)
Age (years)	67 ± 2.5	(46–80)	
Weight (kg)	76 ± 3	(58–101)	
Height (cm)	163 ± 2.5	(149–180)	
BMI	29 ± 1.2	(23.6–37.6)	
Sex (M/F)	6/10		
Anesthesia time (minutes)	204 ± 15	(145–390)	
Surgery time (minutes)	163 ± 16	(110–360)	
Perfusion time (minutes)	91 ± 11	(52–219)	
Reperfusion time (minutes)	30 ± 10	(10–180)	
Aortic clamping time (minutes)	50 ± 4.2	(21–96)	
FVC (%)	94.7 ± 5.0	(37–119)	
FEV_1 (%)	79.0 ± 5.1	(51–109)	
RBC (mL)	576 ± 105	(0–1400)	
FFP (mL)	2 patients received 600 and 800 mL, respectively		
Total fluid balance (mL)	2,763	(1,050–4,800)	
Smoking (Y/N/Ex)?	(3/12/1)		

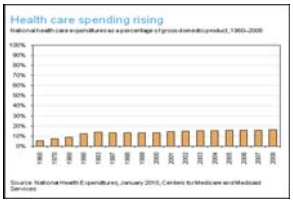
Values are mean ± standard error; (minimum–maximum).

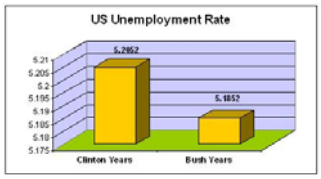
BMI = body mass index; CPB = cardiopulmonary bypass; FEV_1 = forced expiratory volume plasma; FVC = forced vital capacity; OPCAB = off pump coronary artery bypass; RBC = transfusion crystalloid (lactated Ringer's) and hydroxy-ethyl-starch and saline balance.

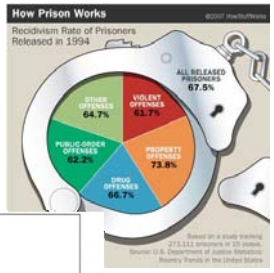
Chart

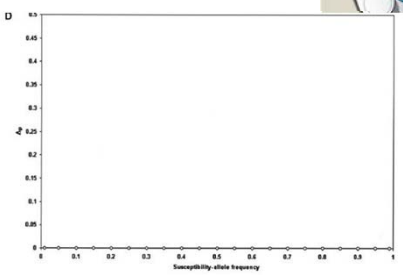


Keep your charts informative rather than decorative!







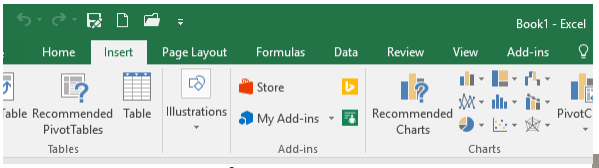


Wittke-Thompson et al, Am J Hum Genet 76: 967-986, Fig 1

"Effective is not the same as beautiful"

Working with charts

- A chart can be embedded in an existing worksheet or can be created on a separate worksheet
- Access: Insert tab/Charts



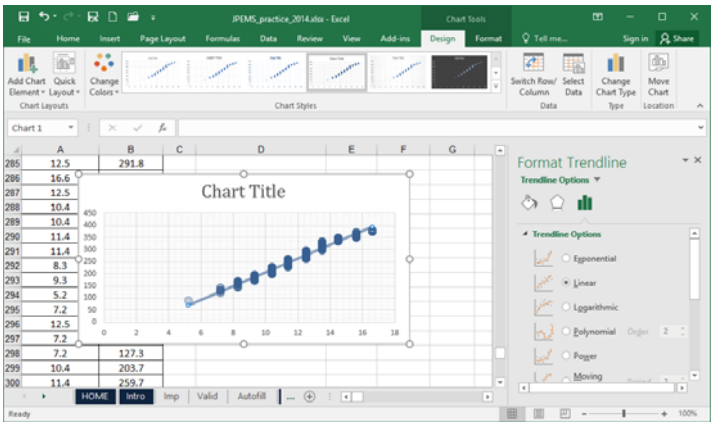
- Steps to create a chart:
  0. Think and plan!
  1. Select the data range (data series)
  2. Select chart type
  3. Insert chart at the specified location
  4. Choose chat options („*Layout*" and „*Design*" tabs)
  5. Change chart style, location, size

Working with charts

Trend and regression analyses if often performed in life sciences

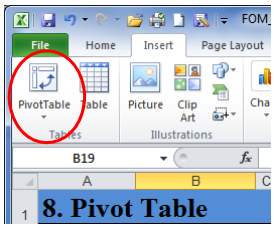
Adding trendline:

- Linear regression
  - Best fit line
  - Regression coefficient
- Curve fitting



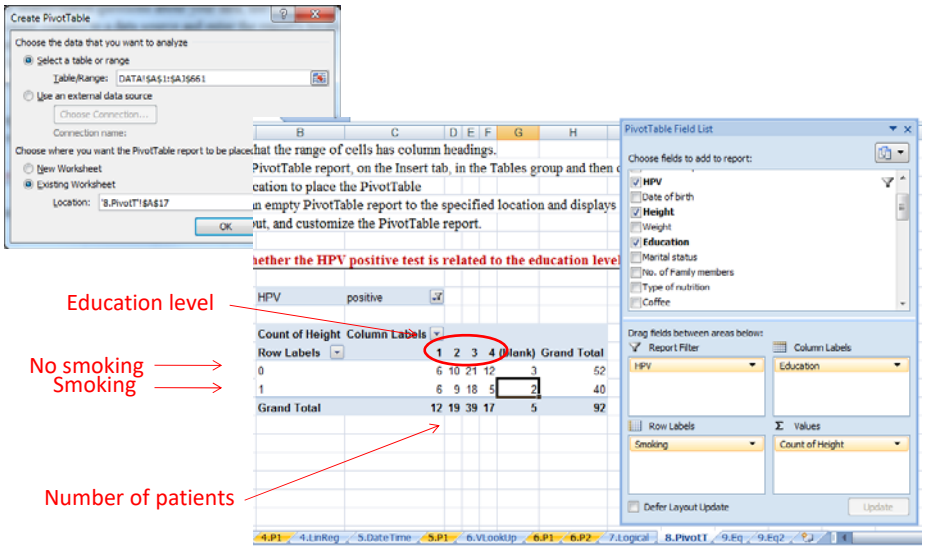
Pivot table

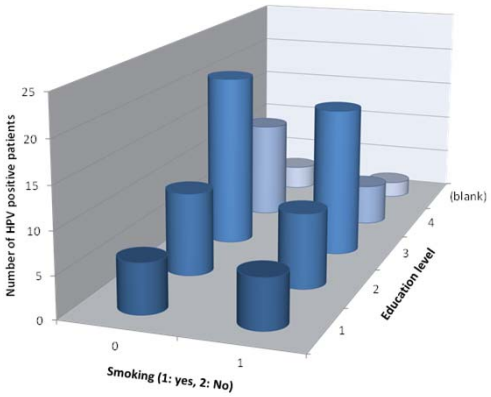
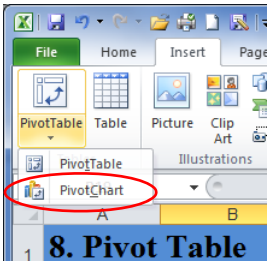
- Summarization tool found in data visualization programs such as **spreadsheets** or business intelligence software.
- To **analyze** numerical **data** in depth and to answer unanticipated questions
  - Lists data in categories
  - Computes summary statistics for those categories
  - Pivot chart to display output



Pivot table - Example

"Is HPV positive test related to the education level or smoking? "



Pivot chart - Example

"Is HPV positive test related to the education level or smoking? "