Prerequisites

- You have installed Tableau Desktop on your computer.

Available here: http://www.tableau.com/academic/students

- You have downloaded the data.

Available here: <u>https://data.nasa.gov/view/angv-aquq</u> (export as .csv) or from our server: <u>http://vda.univie.ac.at/Teaching/Vis/17w/Tutorials/data/</u> <u>Global_Landslide_Catalog_Export.csv</u>

Motivation

With Tableau you can quickly create visualization of your data. Its ease of use makes it valuable tool for initial data exploration and it also allows you to create complex interactive visualizations (e.g. for prototyping).

You can create dashboards consisting of multiple views on the data and options for interaction. Examples are:



Figure 1: http://www.tableau.com/stories/gallery/real-estate-prices



Figure 2: http://public.tableau.com/profile/datalicious.pty.ltd#!/vizhome/ MalteSpitzCallData/MalteSpitzcalldatadashboard



Figure 3: https://public.tableau.com/views/AirbnbSanFranciscoAnalysis/ Airbnb?:embed=y&:loadOrderID=0&:display_count=yes&:showTabs=y

Loading Data

*
Connect
To a file
Excel
Text File
Statistical File
Other Files
To a server
Tableau Server
Microsoft SQL Server
MySQL
Oracle
Amazon Redshift
More Servers

The first step to the visualization is to load the data. On the left side of the screen in the section **To a File** choose **Text File**. This way you can open the .csv file.

After you have selected the data file, you will see an initial table representation of your data. For the data entries to appear, you might have to click **Update Now**.

				Tableau - I	3ook1				
Global_Landslide_Catalog_	Export					Connection		Filters	
Connected to Text File						 Live) Extract	0 Add	
Directory /Users/chris/Desktop	Glot	pal_Landslide_0	Catalog_E						
Files									
Enter file name									
Global Landsg Export.csv									
		Сору				Show a	liases Show hidden	fields Rows 1.693	+
	Id	Date	Time	Country	Nearest Places	Hazard Type	Landslide Type	Trigger	Storm
	#	Ë	Abc	٢	Abc	Abc	Abc	Abc	Abc
	7494	25/09/15	null	null	Barrio Tournon, S	landslide	Landslide	Rain	1
	249	09/09/07	null	Costa Rica	Heredia	landslide	Landslide	Rain	
	250	09/09/07	null	Dominica	entire island, Ros	landslide	Landslide	Rain	Trc
	7541	02/03/16	8:00	null	south bound trave	landslide	Rock_Fall	Unknown	
	7533	27/02/16	12:15	null	Simmons Avenue,	landslide	Rock_Fall	Unknown	
	7423	27/09/15	null	null	Colonia Covias in	landslide	Landslide	Rain	
	6089	23/06/14	null	Nicaragua	El Ayote	landslide	Landslide	Continuous_rain	
	7420	27/09/15	null	null	Colonia La Barran	landslide	Landslide	Rain	
Go to Worksheet ×	6101	23/06/14	null	Nicaragua	El Ayote	landslide	Landslide	Continuous_rain	
🗇 Data Source Sheet 1 🏥 🏥 🚺									

Tableau evaluates automatically what data type is in each column. Always check if the automatic data types are correct by controlling the column headers marked in the figure above.

Global_Landsl	ide_Catalog_E	Mouseover the dataset name and click the little gear to
	Global_Landslide_Catalog_Export.csv	open the settings for
Does the file ind	clude field names in the first row?	the dataset. Choose the correct settings:
No, auto	e first row has field names in it. omatically generate names for the fields.	Field Separator - Comma, tab or other
Field Separator:	Comma 🗘	Text Qualifier - How
Text Qualifier:	н 🗘	can strings be detected
Character Set:	UTF-8	Character Set -
Locale:	English (United States)	© Encoding
<u></u>		Locale -Where is the data from

After you have set the correct format and double checked the column headers, the table is useable. Click **Sheet 1** at the bottom of the window to proceed to your worksheet.

		7420	27/09/15	
	Go to Worksheet	6101	23/06/14	
🗍 Data Source	Sheet 1 🏥 🗄 🕇]		

Basics

Dimensions & Measures

Data Analytics +	In the leftmost panel you will see the columns of your table as either dimensions or measures.
Global_Landshde_Catalo	
Dimensions	Dimensions are usually categorical datatypes. They can be
Abc Adminname1	used to separate your data by discrete tags.
Abc Adminname2	
# Cat Id	
Abc Cat Src	
# Changeset Id	
Abc Comments	
Abc Continentcode	
Country Country	
Countrycode Countrycode	
Countryname	
Date Distance	
Abc. Ceolocation	
Abc Hazard Type	
# Id	
Abc Kev	
Measures	
# Fatalities	Measures are the quantitative data that you will encode in your
# Injuries	marks.
# Population	
# Version	
(1) Latitude (generated)	
Longitude (generated)	You can drag and drop both measures and dimensions to the
*# Number of Records	control panel to create plots
# Measure Values	central parter to create plots.

Building Charts



Let's suppose we want to view the development of the number of landslides over time.

Simply drag the dimension Date to the columns and the measure Number of Records to rows.

You will see that the records are automatically summed up and the date is set to YEAR. Hence the plot shows the number of landslides per year.



We can see here first data quality problems like **null values**. We can remove them by selecting the point and clicking *Exclude*.



Click at the + next to **Year(Date)** to get the next smaller dimension. By clicking on the small triangle next to a Dimension you can change the properties of the dimension. You could also **drag** another Date and change it per hand to month. This allows you to stack an arbitrary amount of dimensions.



The small triangle is available for all datatypes. The measure uses per default **Sum** and you can change it to **Average** or **Variance**. This depends on your dataset and the questions you want to answer/things you want to show.

Hands-On

Question: Which month is the **safest month**? Show the total number of landslides, the number of injured persons and the number of fatalities.

Create a new sheet by clicking on the **New Worksheet** button and answer the question with help of Tableau:

🗍 Data Source	Sheet 1	Sheet 2	ta	ŧ∎	
10 marks 1 row	by 10 column	s SUM(Nu	mbe Ne	w Wo	rksheet 689

If you need help ask the TA's!

One possible solution:



From a total number of events points of view: February is the safest month.

From a number of injured persons points of view: March and April are the safest months.

From a **number of fatalities** point of view: January is the safest month.

Marks & Appearance

Pages	Columns Rows	YEAR(Date) SUM(Number of Re	Back to Sheet 1 . The excluded data is not lost, it is just filtered away. You can always see what filters have been set in the Filters panel.
Filters			
YEAR(Date)	350-		
		1	
Marks Automatic Automatic Color Size Label Detail Tooltip Path	300 - 250 -		Next let's take a look at the Marks panel below it. Here you can modify the appearance of your
	spro200-		plot. We want to use marks to encode more information in view.

Lets find out how many small to very large Landslides happened each year:

• Drag Landslide Size onto Color in the Marks box. We have no multiple lines with weird colors.



•Tableau provides different chart types and by clicking on **Show Me** in the right upper corner you can choose another one. Select **Stacked Bars**. Your view should look something like this:





The default color encoding is actually not that good. Lets change it!

Exclude **Null** and the **2011 Heavy flooding** by right clicking them and select **Exclude**.

Reorder the entries by drag and drop.

We can see here another data quality issue -> upper and lower case answers (e.g. 'Large' and 'large') should not be different groupings

We can fix this by creating a calculated field. To do so, in the **Data Panel**, click the **triangle** right next to **Dimensions.** Click **Create Calculated Field...** and use the following calculation to create a new measure which we will call **clean_size**:

clean_size	\otimes	All	ABS(number)
LOWER([Landslide Size])	Apply OK	Enter Text to Search ABS ACOS AND ASCII ASIN ATAN ATAN ATAN2 ATTR	Returns the absolute value of the given number. Example: ABS(-7) = 7

Drag this new dimension onto **Color**. You should see something like this:



Reorder the entries again by drag and drop.

Click the small triangle which is visible when you mouse over and select Edit Colors...



- -> Assign Palette
- -> Apply
- -> **OK**





It should look like this now!

We can see that most of the Landslides are small or medium.



You can add more information by dragging the **Number of Records** to **Label** in the **Marks Area**.

Hands-On

Question: Which landslide types and how many of them happen in the Dominican Republic, Cuba and Puerto Rico?

Tipps: Use Filters (hint - Quickfilter) and check the dimensions data quality

One possible solution:



Dashboard

A Dashboard connects multiple single views with each other. This allows for more complex questions to be analysed. To create a dashboard click the **new Dashboard** button at the bottom.

Sheet 4	ta	* =	t 1
-)- 01			

Drag Sheets from the left side into the middle area:



You can now define Dashboard wide actions like Filters and Highlights. We will add a filter so that can select in one view data and the second view will be updated accordingly.

In the Dashboard, add a new Action by selecting Dashboard -> Actions...



Add a new action with Add Action > and select Filter...

	Actions [I	Dashboard 1]		
Connect sheets to exterr same workbook using Fi	al web resources (ter actions and Hig	using URL action ghlight actions.	is, or to other sh	eets in the
Name	Run On	Source	Field	s
Add Action >			Edit	Remove
Show actions for all sl	neets in this workb	ook	Cancel	ОК

Name: Filter1 Source Sheets:	Add Filter Acti	on ►	Source Sheets define the sheet where the filter action is recognised. In our case is Sheet 3 the source.
Dashboard 1 Sheet 3 Sheet 4		Run action on: Hover Select Menu Run on single select only	Run action on let you select how the filtering should happen. In our case by selection.
Target Sheets		Clearing the selection will: Leave the filter Show all values Exclude all values	Target Sheets define the sheets where the filter will be applied
Selected Fields	• All Fields Target Field	Target Data Source	Clearing the selection will provides option what happens when you clear the selection.
Add Filter		Edit Remove	

The picture on the right shows the action in action. I have selected medium size landslides in 2014 and we can see that from the three countries we selected in the second hands-on section, only two have valid entries.



Explore

Exercise: Use the mapping functionality of Tableau and create a interactive Dashboard with at least 2 views, to explore the dataset further. Show us what you can find out about triggers, size, type, location, impact and other things!

Lat Abc Loo Abc Loo D Loo Abc Ne	titude cation Accuracy cation Description ngitude ar	Add to Sheet Show Quick Filter Duplicate	bel
Abc Ne Measu	arest Places res	Rename Hide	
# Fat # Inj # Pop	talities uries pulation	Aliases Create	►
# Ve	rsion	Convert to Continuou	IS
 Lai Lai Loi <	titude (generated) ngitude (generated) umber of Records easure Values	Convert to Measure Change Data Type Geographic Role Default Properties	•
 ⊕ Lat ⊕ Lot =# Nu # Me 	titude (generated) ngitude (generated) omber of Records easure Values	Convert to Measure Change Data Type Geographic Role Default Properties Group by Folders Hierarchy	

To use the location data you can use the country information, but there are **null** values in this column. To get **all** data points, you have to **Convert Latitude** and **Longitude** to **Continuous**. After that double click both and the map is ready to go.

If you have problems, ask the Ta's!



A possible solution: Storm Watcher