



FOCUS *lug-ago 2017*

L'USO INIZIALE DI **TABLEAU PUBLIC** PER LA VISUALIZZAZIONE DINAMICA DI DATI

PREMESSA

La visualizzazione grafica e dinamica di informazioni statistiche permette una interazione più efficace fra l'utente e i dati e favorisce la diffusione delle informazioni statistiche, obiettivo, tra gli altri, degli Uffici di Statistica. In questo modo, si può offrire una vista attiva dei dati, agevolando le speculazioni e le analisi che si realizzano attraverso la selezione dinamica delle variabili e della tipologia di grafico da rappresentare, ad esempio una mappa piuttosto che un diagramma. L'ambito tecnologico che ha sviluppato le soluzioni informatiche in grado di rispondere a questo bisogno prende il nome di *Business Intelligence* (BI). Fra i numerosi *tool* per la visualizzazione dinamica, l'Ufficio Statistico della regione Puglia ha scelto di utilizzare *Tableau Public*, opzione *free* dell'offerta commerciale *Tableau*, che vincola l'utilizzatore a salvare i propri grafici su un *cloud* remoto, rendendoli accessibili alla collettività e in linea con la logica *open* della Pubblica Amministrazione. In questo FOCUS si ha l'obiettivo di condividere l'esperienza operativa nell'uso iniziale di questo strumento,

prendendo come riferimento i dati ISTAT sui sistemi locali del lavoro (SSL).

Fonti dei dati: [Istat – Sistemi Locali del lavoro nel 2011](#).

Cosa è Tableau

Tableau Public è il servizio *free*, offerto dalla società *Tableau*, per la costruzione, la visualizzazione e la pubblicazione *on line* di tabelle, grafici e mappe dinamiche interattive. Per poter usufruire di tale servizio è necessario: 1) [scaricare il software sul proprio PC](#), inserendo un indirizzo di posta elettronica valido, prima di attivare il processo di *downloading*; 2) installare il *software*; 3) crearsi un *account* sul *cloud* di *Tableau*. Il *software* permette la connessione dei dati che risiedono in locale ai grafici sul *cloud*, la creazione delle visualizzazioni interattive e il salvataggio degli elaborati sul *cloud* remoto. Attraverso un insieme di istruzioni *HTML* oppure un indirizzo web, gli oggetti grafici del *cloud* possono essere inglobati in pagine *web* o condivisi su piattaforme *social* o diventare collegamenti ipertestuali in documenti. Questi oggetti sono comunque pubblici e quindi visibili a tutti. Non è permesso il salvataggio degli elaborati su memorie locali. Per rispondere anche alle altre diverse e più complesse esigenze aziendali di BI, quali le decisioni strategiche legate alla trasformazione di dati e di informazioni in conoscenza, *Tableau* offre una serie di prodotti commerciali. La soluzione *Tableau Public* è in accordo con il tema della garanzia, trasparenza e partecipazione dei cittadini alla *res publica* che si realizza anche attraverso l'*open data*, descritto dal Codice dell'amministrazione digitale come uno degli elementi di innovazione della Pubblica Amministrazione (PA). L'obiettivo è rendere i dati della PA fruibili e accessibili liberamente. Il servizio *Tableau Public* può essere inteso, in tal senso, come uno strumento di valorizzazione degli *open data*.

Nella visualizzazione dinamica con *Tableau Public* è necessario seguire le seguenti fasi: 1) preparazione dei dati; 2) connessione ai dati; 3) costruzione *sheet*; 4) predisposizione

dashboard; 5) organizzazione delle story; 6) pubblicazione *on line* dei grafici.

Preparazione dei dati

Per scaricare il software *Tableau Public* è necessario fornire una *e-mail* valida e crearsi un *account* sul *cloud* di *Tableau*, che permette di gestire lo spazio destinato all'utente. Il programma si installa agevolmente sul proprio PC.

Nell'esempio che segue si sono selezionati alcuni [dati sui SLL da fonte Istat](#) e i relativi [shape file](#) che descrivono i loro confini poligonal. I SLL, nell'accezione proposta dall'Istat fin dal 1981, rappresentano l'unione di comuni all'interno dei quali la popolazione risiede e lavora ed esercita la maggior parte delle proprie relazioni sociali ed economiche.

I principali formati accettati in *input* dal programma sono le tabelle *Excel* oppure i *file comma separated value*, quelli con estensione *.csv*, composti da testo semplice separato da virgole. Le informazioni possono essere contenute in uno o più fogli di lavoro *Excel*. Nella prima riga di dati si consiglia di inserire i nomi delle variabili.

Il programma può leggere i dati geografici nei vari formati (*ESRI shapefile*, *MapInfo tables*, *KML files*, *GeoJSON files*). Si poggia sulla tecnologia libera *OpenStreetMap*.

In ciascun progetto, i dati possono essere contenuti in più file di vario formato e in più fogli. E' necessario organizzarli per unità di informazione. Nel caso in esame, si tratta dei SLL, identificati dal codice Istat assegnato dopo il censimento del 2011 (COD_SLL_2011_T).

Il primo file "[tassi SSL.xlsx](#)", oltre all'identificativo del SLL, contiene i seguenti dati: la denominazione del SLL (denominazione), il codice Istat della regione di appartenenza (REG), il tasso di attività, di occupazione e di disoccupazione relativi all'anno 2016. Le prime tre righe del file sono in fig. 1.

Fig.1 - Prime righe del file *tassi SSL.xlsx*

COD_SLL_2011_T	Denominazione	REG	Tasso di attività	Tasso di occupazione	Tasso di disoccupazione
0101	CHIERI	1	52,8	48,2	8,7
0102	IVREA	1	51,2	46,5	9,3
0103	PINEROLO	1	51,3	46,9	8,7

Il campo COD_SLL_2011_T è stato costruito in formato testo, inserendo lo zero davanti ai codici di sole tre cifre, in modo da renderlo compatibile con il campo dello *shape file* dei sistemi locali (SLL_2011_T).

Il codice REG può essere abbinato alla denominazione della regione attraverso il campo *cod_regioni* del file "[cod_regioni.xlsx](#)" (tab. 1).

Tab. 1 - Tabella di corrispondenza fra codici Istat e denominazione delle regioni (file *cod_regioni.xlsx*)

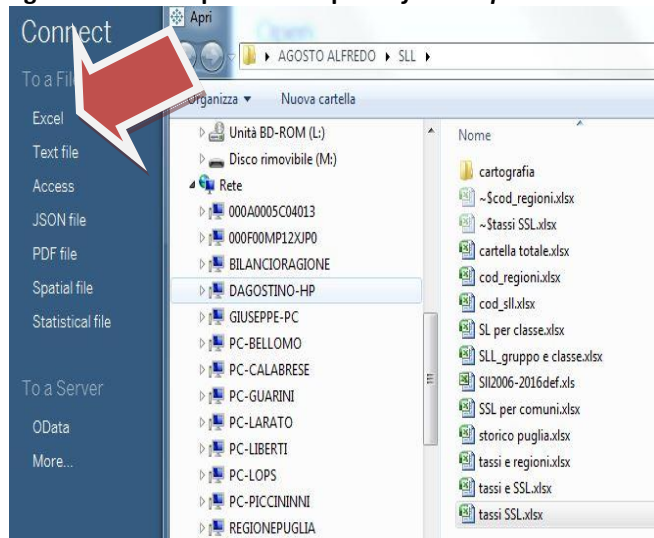
cod_regioni	Regioni
1	Piemonte
2	Valle d'Aosta
3	Lombardia
4	Trentino Alto Adige
5	Veneto
6	Friuli-Venezia Giulia
7	Liguria
8	Emilia-Romagna
9	Toscana
10	Umbria
11	Marche
12	Lazio
13	Abruzzo
14	Molise
15	Campania
16	Puglia
17	Basilicata
18	Calabria
19	Sicilia
20	Sardegna

La *shape file* "[SLL2011 wgs84 EURO.shp](#)" dei SLL va inserito in una *directory*, insieme a tutti gli altri *file* geografici, quelli con estensione [.shx](#) e [.dbf](#).

Connessione ai dati

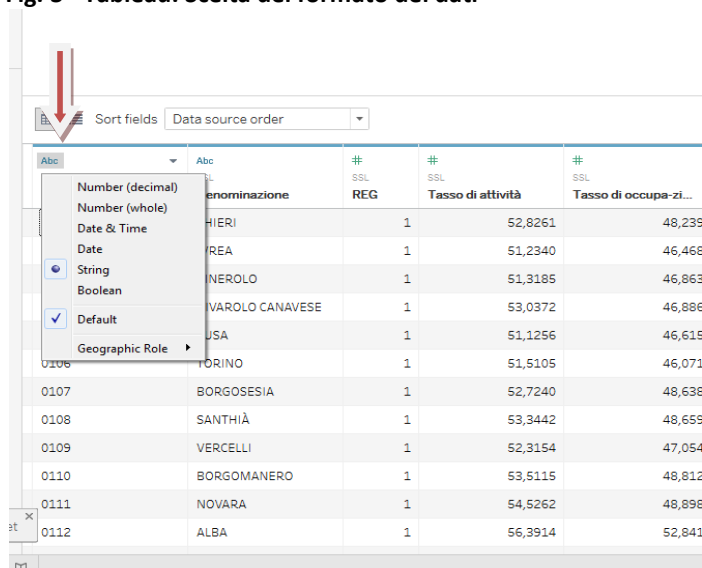
Lanciato *Tableau*, dal menù *Connect* si può selezionare la tipologia di file dati da elaborare. Nel caso in esame, due sono *Excel* e uno è *shape* o *Spatial file*. Cliccando su *Excel* e sfogliando le *directory* locali si sceglie e si apre il primo file d'interesse (*tassi SSL.xlsx*), come si vede dalla fig. 2.

Fig. 2 - Tableau. Apertura del primo file in input



I dati sono letti, interpretati da *Tableau* rispetto al formato e presentati in forma di tabella. Nel caso si debba modificare un formato, si può cliccare sul simbolo "Abc" (alfanumerico), oppure "#" (numerico) e scegliere quello più appropriato dal menù a cascata (fig. 3)

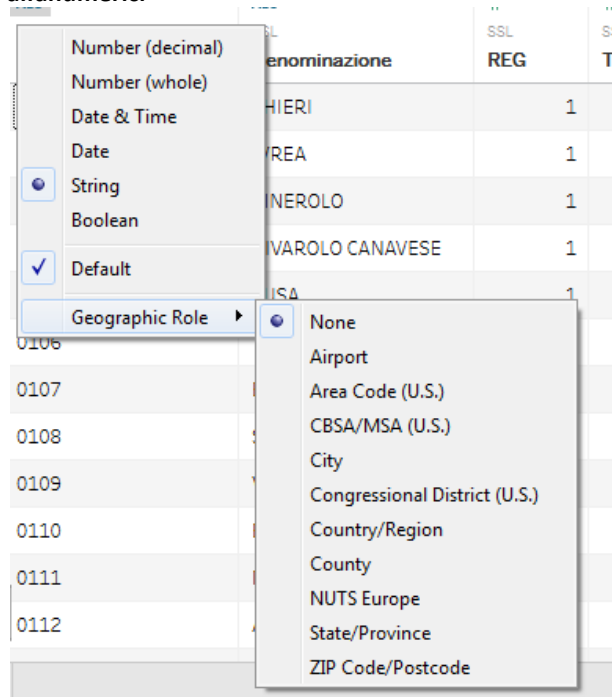
Fig. 3 - Tableau. Scelta del formato dei dati



Grazie alla tecnologia *OpenStreetMap*, un dato può essere associato a confini amministrativi

assegnandogli un ruolo geografico, come nel caso in cui una colonna contenga i nomi delle regioni a cui venga associato il ruolo "NUTS Europe" (fig. 4).

Fig. 4 - Tableau. Scelta del ruolo geografico per i dati alfanumerici



Un campo numerico può assumere il ruolo geografico di coordinata "latitude" oppure "longitude" (fig. 5).

Fig. 5 - Tableau. Scelta del ruolo geografico per i dati numerici

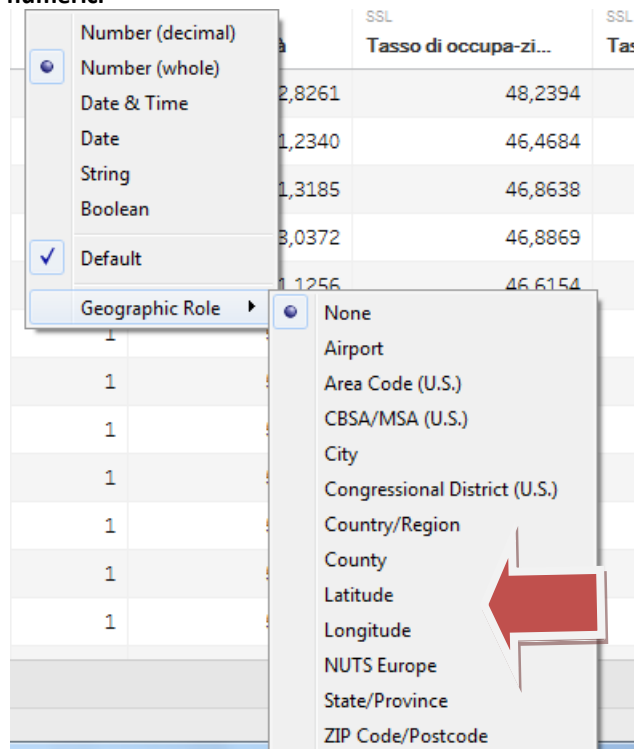
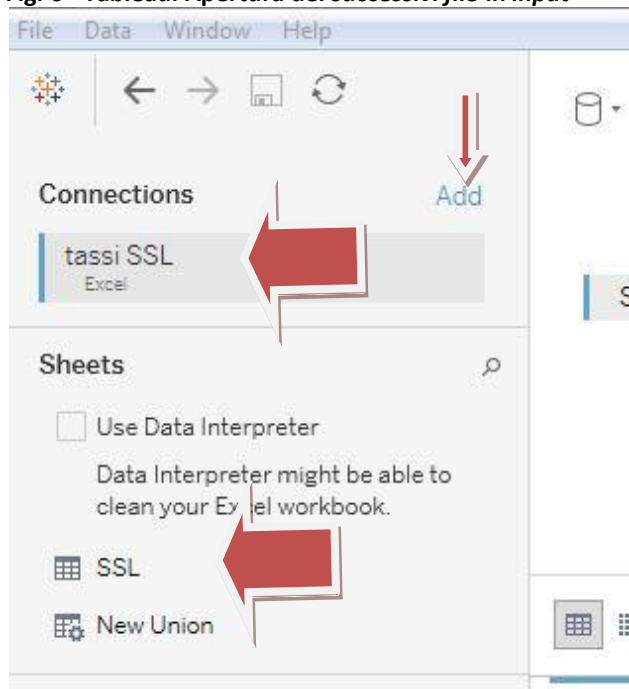


Tableau mostra le connessioni ai file e i nomi dei fogli già aperti. Per aggiungere altri dati è necessario cliccare su “Add”.

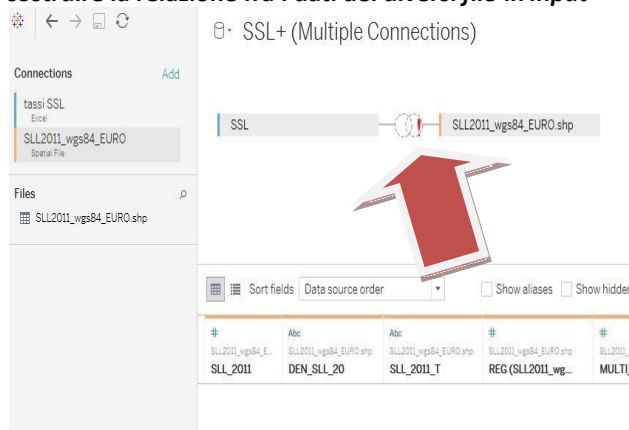
Fig. 6 - Tableau. Apertura dei successivi file in input



Per connettere lo *shape file* si deve selezionare la opzione *Spatial file*. Anche in questo caso, si deve selezionare la *directory* contenente i file geografici e puntare su quello con estensione *.shp* (*SLL2011_wgs84_EURO.shp*).

Il punto esclamativo di colore rosso avverte che è necessario individuare un campo di collegamento fra i *file*, in modo che si possa creare la congiunzione dei dati (fig. 7).

Fig. 7 - Tableau. Avvertimento sulla necessità di costruire la relazione fra i dati dei diversi file in input

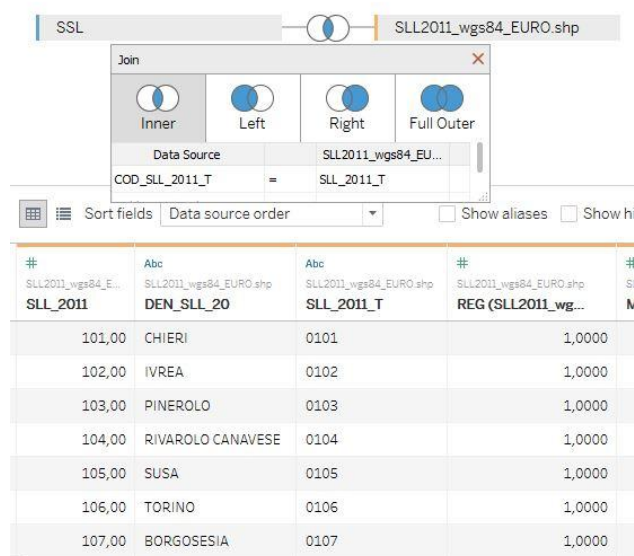


Cliccando sul punto esclamativo si apre una finestra da cui scegliere i due campi. Dal lato

Data Source si può aprire il menù a tendina per scegliere il primo codice (*COD_SLL_2011_T*); sull'altro lato è presente il nome del file geografico, sotto il quale si apre la tendina per scegliere il campo di collegamento (*SLL_2011_T*). La colorazione *blu* indica la coerenza dei campi prescelti. I due file hanno così messo in comune tutte le informazioni sull'unità di informazione SLL (fig. 8).

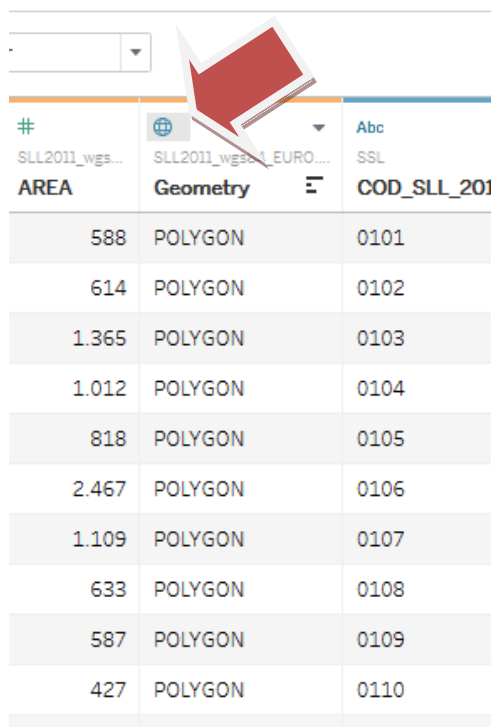
Fig. 8 - Tableau. Scelta delle variabili per la costruzione della relazione fra i dati dei diversi file in input

SSL+ (Multiple Connections)



Scorrendo tutte le colonne dei dati, si scopre che il ruolo geografico è stato associato automaticamente alla variabile *Geometry* dello *shape file* ed identificato dal simbolo del mappamondo (fig. 9).

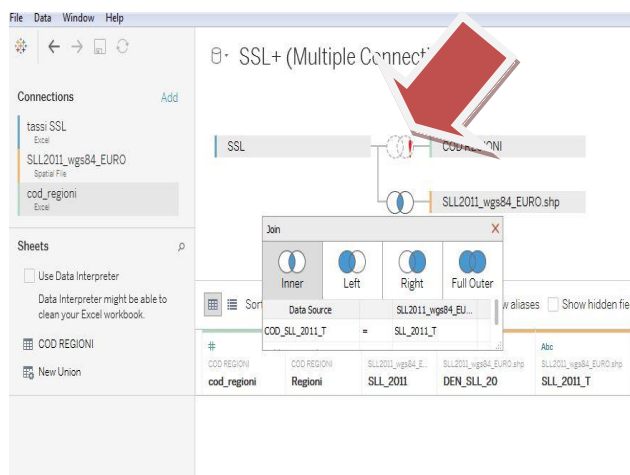
Fig. 9 - Tableau. Individuazione ruolo geografico all'interno delle variabili dello *shape file*.



#	AREA	Geometry	COD_SLL_201
588	POLYGON		0101
614	POLYGON		0102
1.365	POLYGON		0103
1.012	POLYGON		0104
818	POLYGON		0105
2.467	POLYGON		0106
1.109	POLYGON		0107
633	POLYGON		0108
587	POLYGON		0109
427	POLYGON		0110

L'ultimo file da connettere è quello dei codici-nomi regione. Anche in questo caso, si tratta di cliccare su "add" e "connect To a File->Excel" e sfogliare il percorso per raggiungere il file "*cod_regioni.xlsx*". Effettuata la scelta e aperto il file, ci si trova nella situazione precedente: l'esigenza, cioè, di collegare i dati fra loro (fig. 10).

Fig. 10 - Tableau. Avvertimento sulla necessità di costruire la relazione fra i dati dei diversi *file in input*

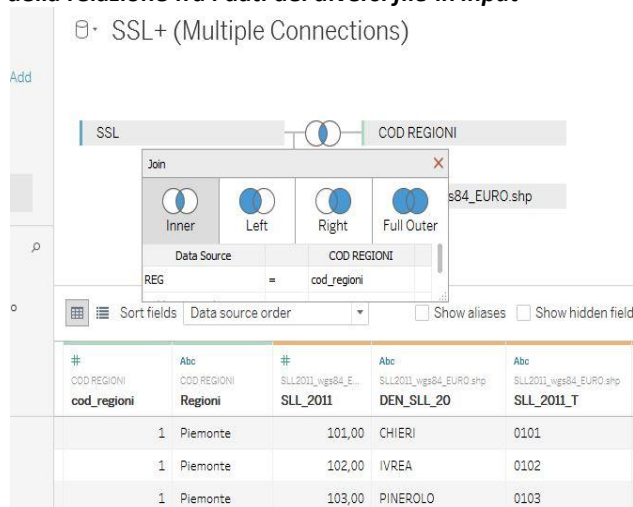


#	COD REGIONI	COD REGIONI	SLL2011_wgs84_E...	SLL2011_wgs84_EURO.shp	SLL2011_wgs84_EURO.shp
	cod_regioni	Regioni	SLL_2011	DEN_SLL_20	SLL_2011_T
1	Piemonte		101,00	CHIERI	0101
1	Piemonte		102,00	IVREA	0102
1	Piemonte		103,00	PINEROLO	0103

Cliccando sul punto esclamativo, dalla tendina del lato *Data Source*, si sceglie il campo REG, da

quella del lato COD REGIONE, si scegli il campo *cod_regioni* (fig. 11).

Fig. 11 - Tableau. Scelta delle variabili per la costruzione della relazione fra i dati dei diversi *file in input*



#	COD REGIONI	COD REGIONI	SLL2011_wgs84_E...	SLL2011_wgs84_EURO.shp	SLL2011_wgs84_EURO.shp
	cod_regioni	Regioni	SLL_2011	DEN_SLL_20	SLL_2011_T
1	Piemonte		101,00	CHIERI	0101
1	Piemonte		102,00	IVREA	0102
1	Piemonte		103,00	PINEROLO	0103

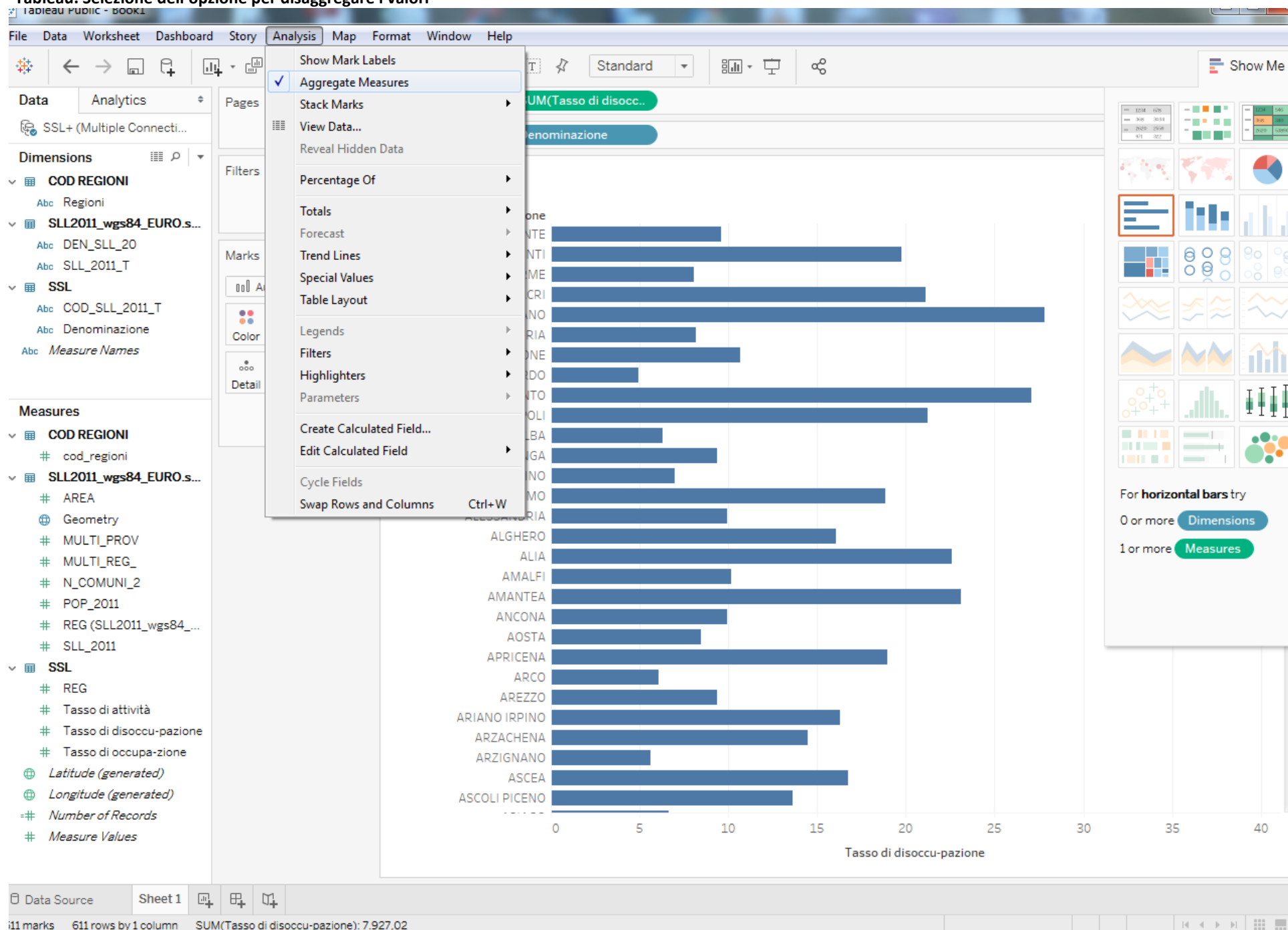
I dati sono tutti collegati fra loro e pronti per le successive elaborazioni.

Costruzione *Sheet*

Nella logica di Tableau, una storia (*Story*) può contenere due o più cruscotti (*Dashboard*); ciascun cruscotto è costruito sulla base di un grafico, contenuto all'interno di un foglio (*sheet*). Si parte, quindi, dalla preparazione di uno *sheet*.

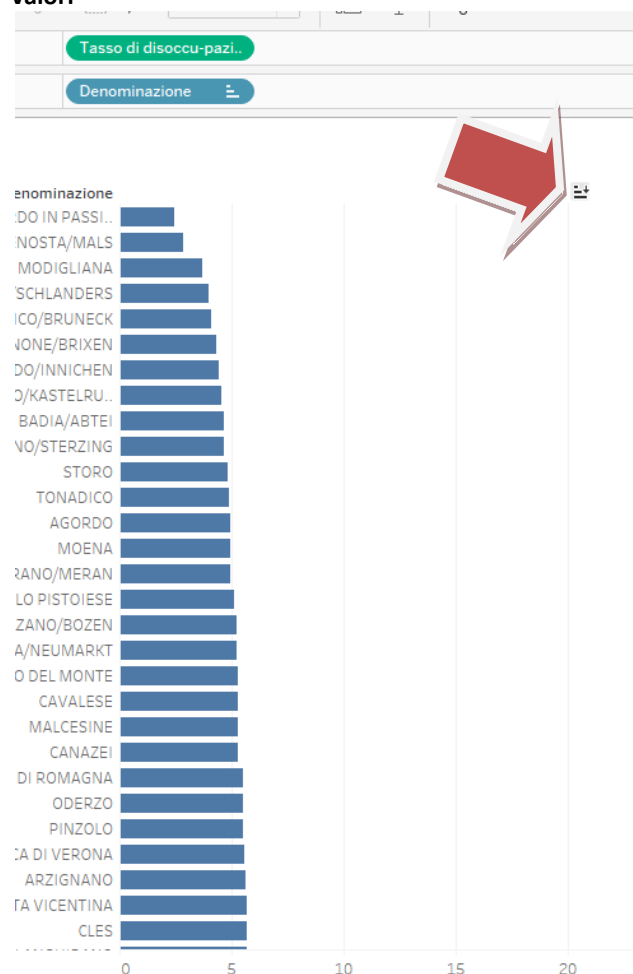
Con la finalità di rappresentare attraverso un diagramma a barre il tasso di disoccupazione dei SSL in Italia, si trascina nello spazio "*Columns*" la variabile "tasso di disoccupazione" e in quello "*Rows*" la variabile "Denominazione". Bisogna porre attenzione alla opzione "Aggregate Measures" del menù "Analysis". Inizialmente questa si presenta selezionata e, quindi, le variabili sono sommate. Togliendo la spunta, le variabili saranno considerate singolarmente ed è questa la condizione con cui si prosegue l'esempio (fig. 12).

Fig. 12 – Tableau. Selezione dell'opzione per disaggregare i valori



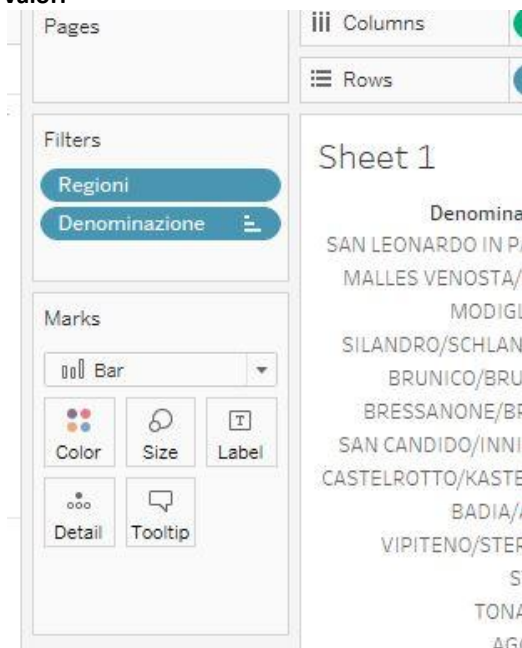
Attraverso il simbolo evidenziato dalla freccia rossa in fig. 13 è possibile ottenere l'ordinamento crescente o decrescente dei SLL, in funzione del tasso di disoccupazione.

Fig. 13 – Tableau. Selezione dell'opzione per ordinare i valori



Il diagramma può diventare dinamico ed interattivo se si fornisce all'utente la possibilità, ad esempio, di selezionare il singolo SLL o tutti quelli appartenenti ad una data regione. Lo spazio "Filters" risponde a questa esigenza. Si trascina, quindi, il campo "Regioni" e quello "Denominazione" in questo spazio (fig. 14).

Fig. 14 – Tableau. Selezione dell'opzione per filtrare i valori

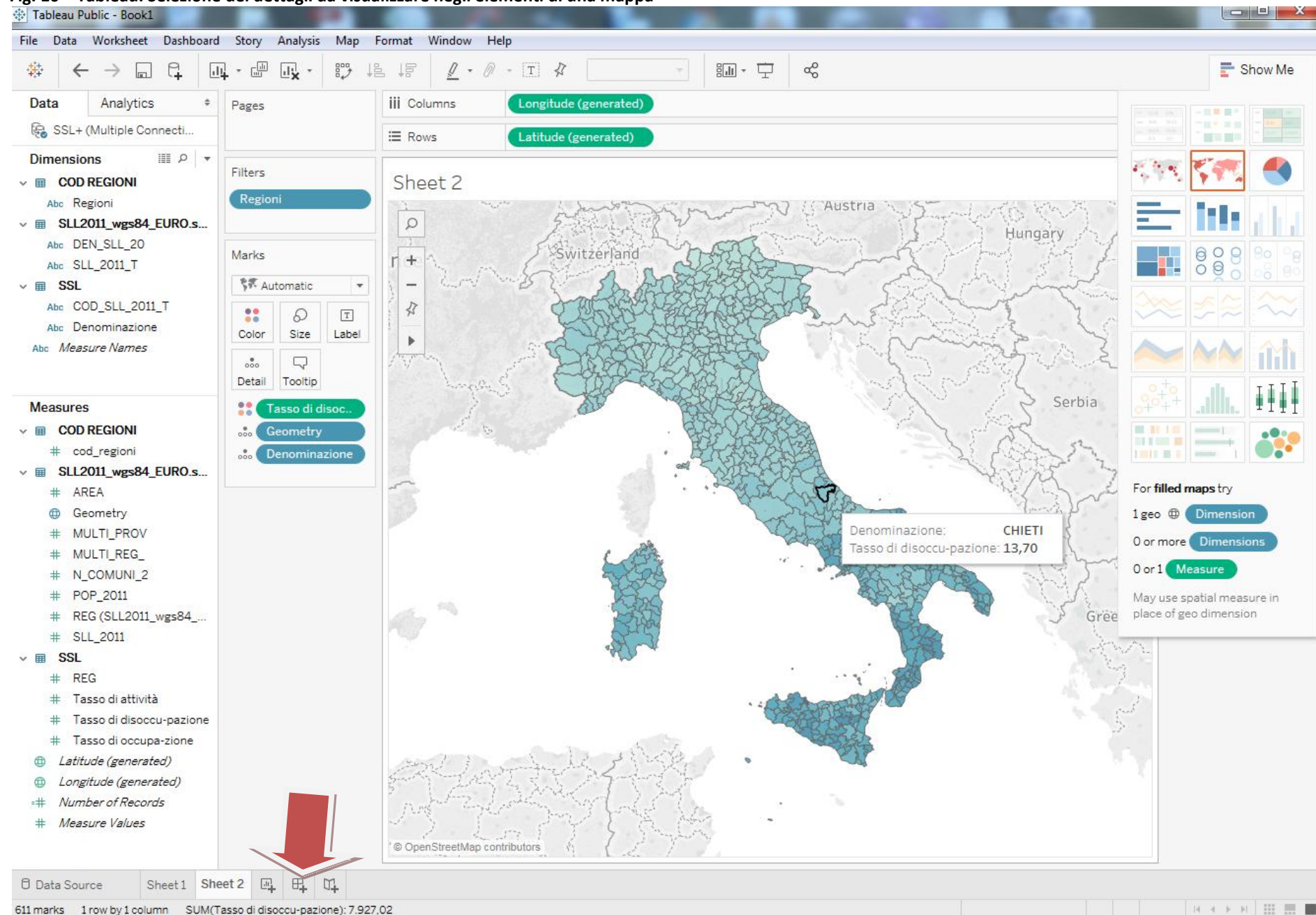


Al fine di costruire una mappa dei SLL si seleziona un nuovo foglio. Cliccando due volte sulla variabile che gioca il ruolo geografico, nel caso in esame "Geometry", dopo pochissimi secondi appare a video una mappa con tutti i poligoni che identificano i SLL (fig. 15). La mappa diventa tematica, trascinando una variabile numerica all'interno del marcatore colore (Marks->Color), come indicato nella freccia di fig. 15. Diventa dinamica, trascinando nella sezione "Filters" (vedi freccia di fig. 15) il campo prescelto ("Regioni"). In fig. 16 si presenta la mappa del tasso di disoccupazione, con la possibilità di selezionare tutti i SLL di una data regione. Nell'esempio si è aggiunto il campo "Denominazione" nel quadrato "Details" della sezione "Marks". In questo modo, spostando il mouse sulla mappa e cliccandoci sopra si seleziona il SLL e si apre un riquadro con le informazioni presenti nei "Details", come si evince dalla stessa fig. 16.

Predisposizione dashboard

Un primo oggetto grafico pubblicabile sul web è il cruscotto o *dashboard*. Nell'esempio, si produrranno due *dashboard*, uno per ciascuno *sheet*. Per selezionare l'opzione è necessario cliccare sull'icona, indicata dalla freccia rossa in fig. 16.

Fig. 16 – Tableau. Selezione dei dettagli da visualizzare negli elementi di una mappa



Il *dashboard* si presenta come in fig. 17. Sul lato sinistro sono indicati i nomi dei fogli. Si seleziona *Sheet1*, trascinandolo nel cruscotto. Al diagramma si possono aggiungere una serie di opzioni dinamiche. Cliccando sul triangolo in alto a destra (fig. 17) si ha la possibilità di accedere ad un menù e a vari sotto menù, tra i quali quello della selezione dei filtri (Denominazione, Regioni e Tasso di disoccupazione). Si sceglie "Regioni". Operando sul triangolo in alto a destra, evidenziato sempre con la freccia in fig. 18, si accede ad un ulteriore menù attraverso il quale si decide il tipo e la modalità di selezione del filtro, ad esempio, "a cascata" e più regioni. Si ripete l'operazione con tutti gli altri filtri che si intende inserire nel cruscotto. Il filtro numerico, ad esempio quello relativo ai tassi di disoccupazione, permette di selezionare i SLL sulla base di intervalli di valori dei tassi. Cliccando due volte nel box "*Sheet*" del cruscotto è possibile assegnare un titolo al cruscotto. Tutti i nomi degli elementi grafici presenti nel *dashboard* possono essere cambiati col doppio click (fig. 19). Il secondo *dashboard* è costruito sulla base del foglio *Sheet2*, seguendo gli stessi criteri precedentemente esposti (fig. 20).

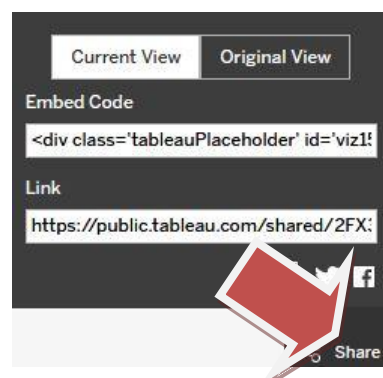
Organizzazione delle *story*

La modalità di rappresentazione *story* permette di unire più *dashboard*.

Una volta cliccato sull'icona *story* del menù in basso, si ha la possibilità di trascinare il primo cruscotto nella *story*, specificando il contenuto all'interno del rettangolo di selezione contenuto tra i due simboli "<>". Il secondo *dashboard* può essere trascinato affianco al primo rettangolo: in automatico si aprirà una nuova finestra contenente l'oggetto grafico, nel nostro caso la mappa. I due tasti "Diagramma" e "Mappa" permettono di selezionare il cruscotto e interrogare i dati (fig. 21) sulla base dei filtri. Nella fig. 22, dopo aver selezionato il bottone "Mappa", si fornisce un esempio di possibile interrogazione dei dati.

Pubblicazione *on line*

Per pubblicare *on line* un cruscotto oppure una *story* è necessario selezionare il menù "File", l'opzione "Save to Tableau Public as". Sarà richiesto l'account di registrazione sul *cloud* di Tableau e il nome da assegnare al progetto (*Workbook*). Dopo il salvataggio, cliccando sul tasto "share", l'oggetto può essere condiviso



su una pagina web, copiando il codice HTML messo a disposizione da Tableau (*Embed Code*), oppure attraverso un link. Cliccando, invece, sul tasto

"download" è possibile scaricare quattro

differenti tipi di file: il pdf dell'oggetto grafico; una tabella d'incrocio delle variabili (*crosstab*); il file testo di tutti i dati del progetto; il file in formato immagine (*jpg*) del grafico.



Il seguente [link](#) permette l'accesso alla *story* introdotta in questo Focus.

Dal box "Diagramma" si accede al grafico a barre orizzontali del tasso di disoccupazione 2016. A livello nazionale, San Leonardo in Passiria (2,44) e Malles Venosta (2,85) della Provincia Autonoma di Bolzano sono i SLL con il minore tasso di disoccupazione; Bagheria (39,08) in Sicilia e Rosarno (31,38) in Calabria sono quelli con il tasso più alto. Corrispondentemente, in Puglia, si ha Ostuni (14,56) e Martina Franca (14,82) con i valori più bassi e Cerignola (26, 42) e Ugento (26,07) con quelli più alti.

Dal box "Mappa" si accede alla rappresentazione cartografica del tasso di disoccupazione sull'intero territorio nazionale.

Sulla scala di colori è associato l'intervallo di valori compreso fra 2,44 e 39,08. Si evidenziano colori più tenui nelle regioni del nord, più scuri nei SLL delle regioni meridionali dove il fenomeno è più rilevante.

Fig. 17 – Tableau. Selezione dei filtri di un cruscotto

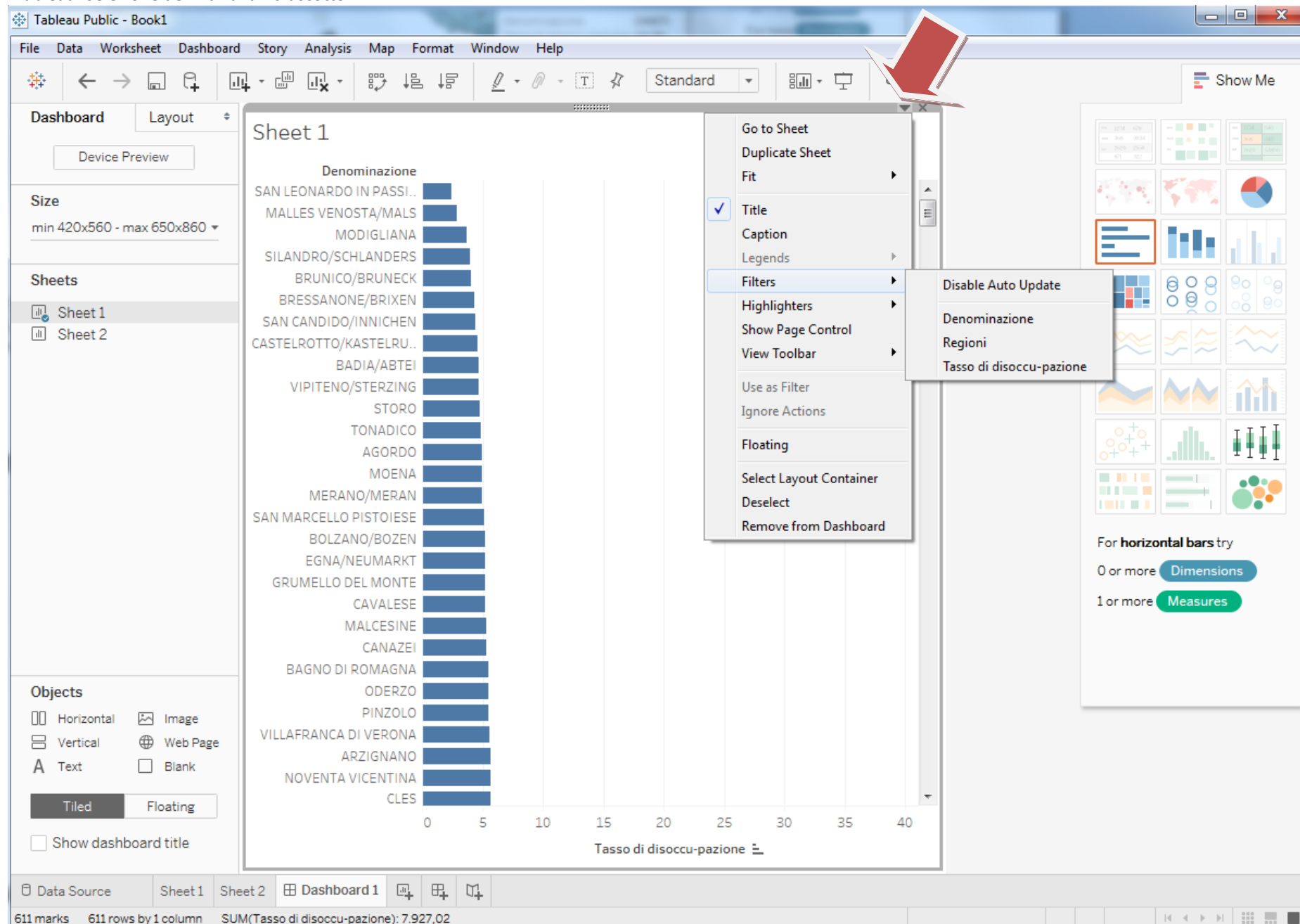


Fig. 18 – Tableau. Selezione della tipologia di formato per i filtri di un cruscotto

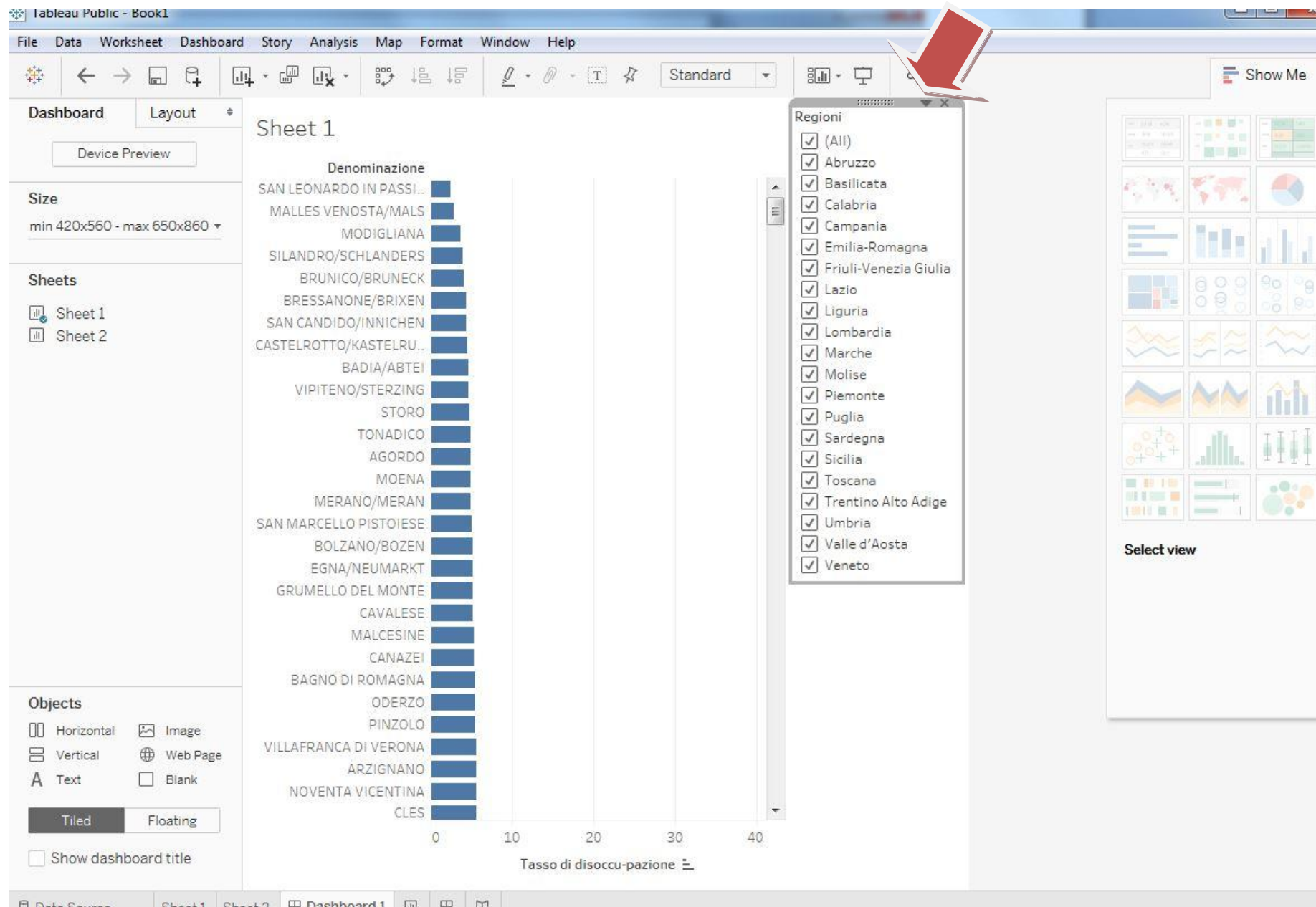


Fig. 19 – Tableau. Selezione della regione dal filtro di un cruscotto

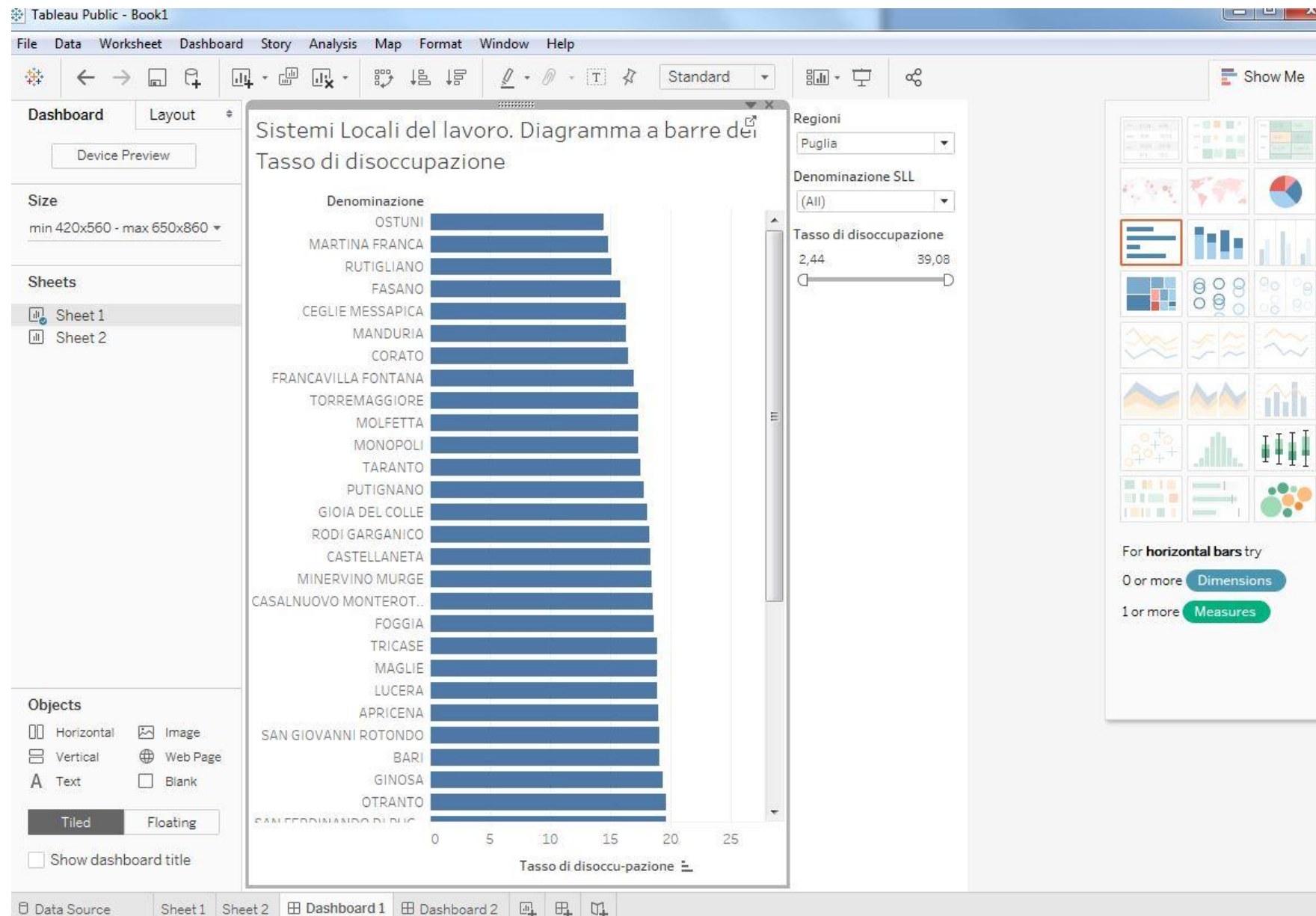


Fig. 20 – Tableau. Costruzione di un cruscotto contenente una mappa

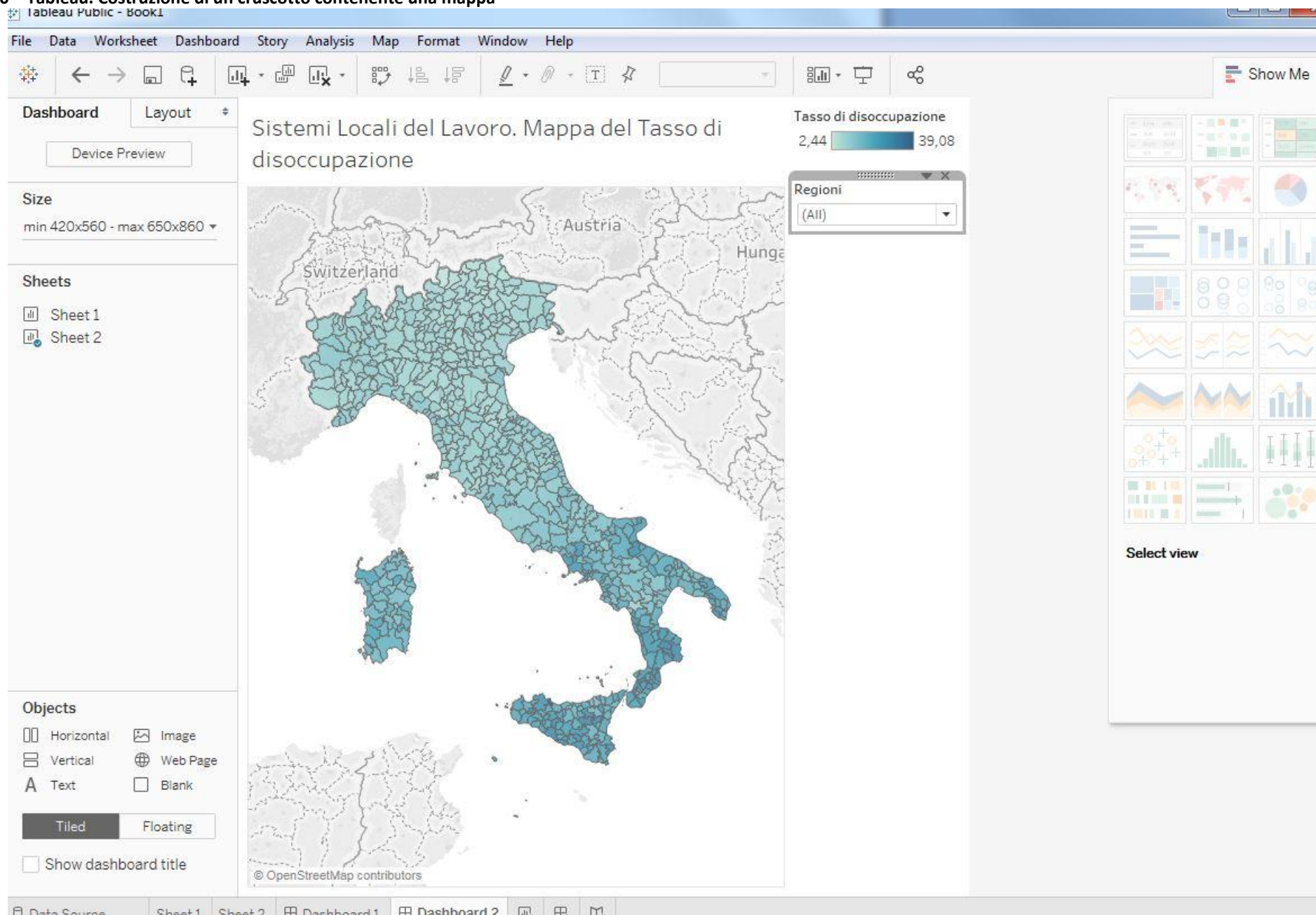


Fig. 21 – Tableau. Costruzione di una story

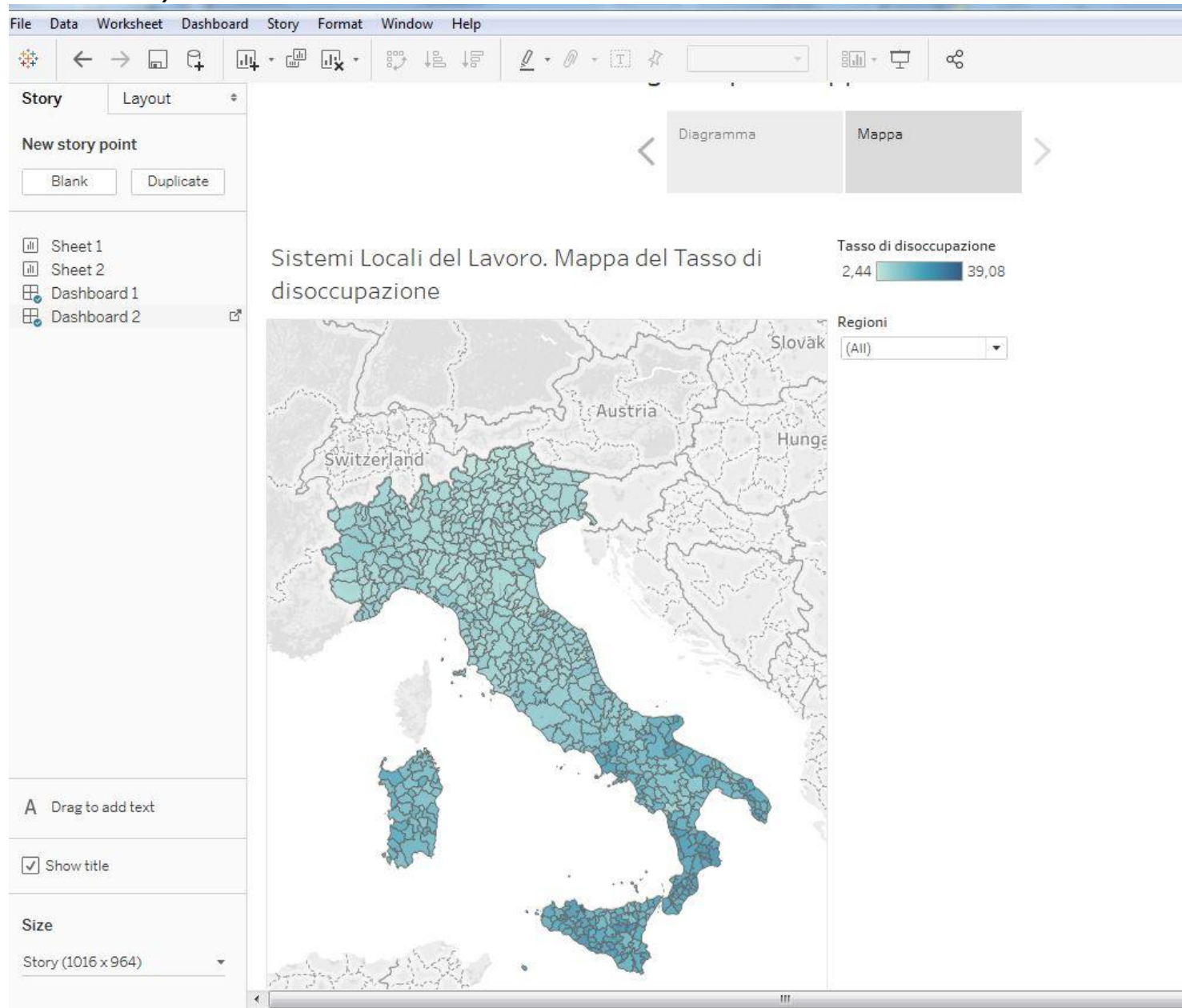
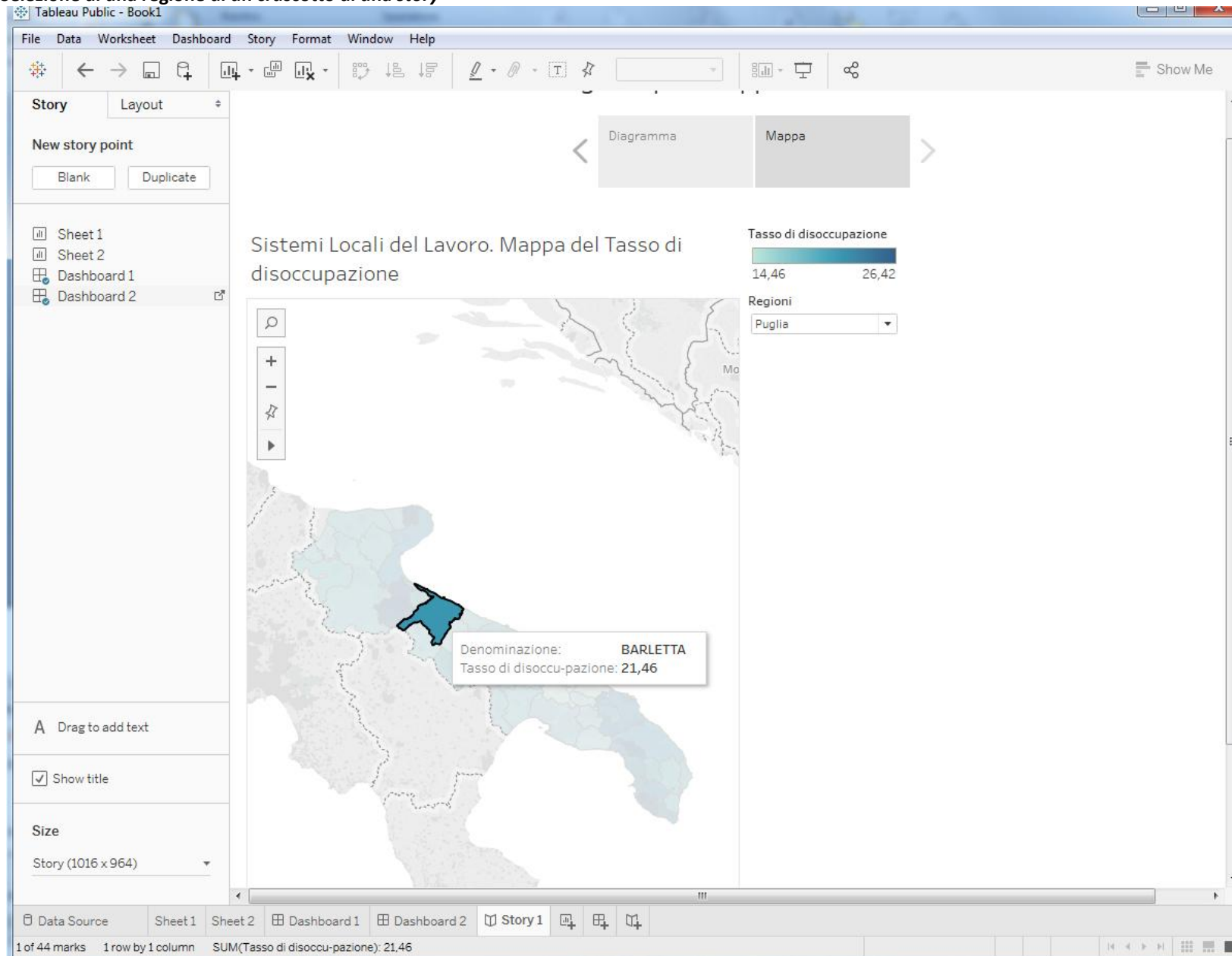


Fig. 22 - Tableau. Selezione di una regione di un cruscotto di una story



Glossario

Sistema locale del lavoro - Unità territoriale identificata da un insieme di comuni contigui legati fra loro dai flussi di pendolarismo. I sistemi locali ripartiscono esaustivamente il territorio nazionale, prescindendo da altre classificazioni amministrative. Consentono la diffusione di informazione statistica su una base geografica di aree funzionali. Sotto il profilo metodologico i sistemi locali sono costruiti come aggregazione di comuni che soddisfano requisiti di dimensione (almeno 1.000 occupati residenti) e di livelli minimi d'interazione espressi tramite funzioni di auto-contenimento.

Business Intelligence - Sistema di modelli, metodi, processi, persone e strumenti che rendono possibile la raccolta regolare ed organizzata del patrimonio dati generato da un'azienda. Inoltre attraverso elaborazioni, analisi o aggregazioni, ne permettono la trasformazione in informazioni, la loro conservazione, reperibilità e presentazione in una forma semplice, flessibile ed efficace, tale da costituire un supporto alle decisioni strategiche, tattiche ed operative. Il sistema di Business Intelligence comporta dunque: la raccolta dei dati del patrimonio dell'azienda, la loro pulizia, validazione e integrazione, la successiva elaborazione, aggregazione e analisi l'utilizzo fondamentale di questa mole di informazioni nei processi decisionali

OpenStreetMap – E' un'iniziativa aperta alla creazione dei dati geografici ed alla loro libera circolazione, senza oneri e verso chiunque li desideri.

UFFICIO STATISTICO, Via Gentile 52 - 70126 Bari
email: ufficio.statistico@regione.puglia.it
www.regione.puglia.it/web/ufficiostatistico