第5回（2019年10月18日）

オープンデータから横断データ集を作って活用しよう

株式会社ウェルネス
渡部　鉄兵
https://www.wellness.co.jp/

こんなことありませんか …

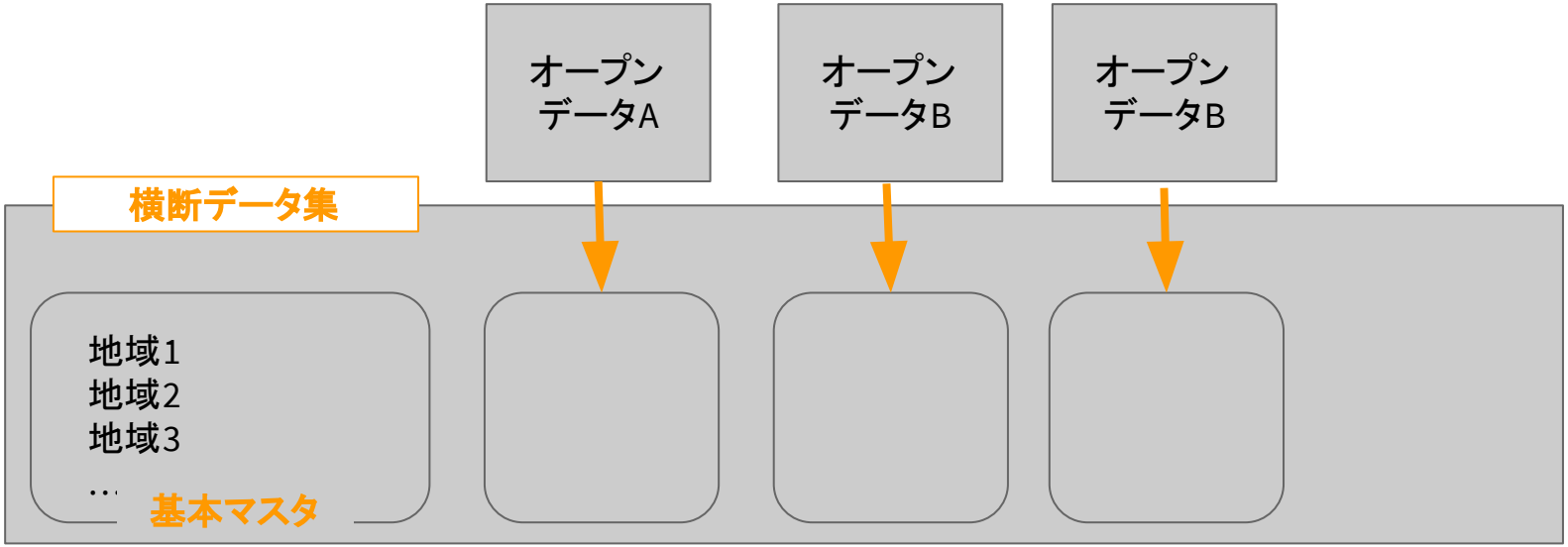
- オープンデータひとつひとつでは関連性がみえてこない…
- 既存のデータ集では物足りない…

オープンデータを組み合わせて、自分だけの横断データ集を作ろう

本日の要旨

- データ集の基本となるマスタを作ろう
- 連結したいオープンデータに下処理をしよう（初級編）
- マスタにオープンデータを連結しよう
- 連結したいオープンデータに下処理をしよう（中級編）
- 横断データ集を完成させよう

横断データ集のイメージ



基本マスタ

データ集の基礎となる地域単位

- 国
- 都道府県
- 二次医療圏
- 市区町村
- ・・・

「二次医療圏」単位で集計されたオープンデータはそれほど多くはない！

→ 必要に応じて市区町村データから集計する必要があります（中級編）

基本マスタ：二次医療圏マスタを作ろう

e-Stat 政府統計の総合窓口
https://www.e-stat.go.jp/

基本マスタ：二次医療圏マスタを作ろう

データセット一覧

<戻る

政府統計名	医療施設調査
提供統計名	医療施設調査
提供分類1	平成２９年医療施設（静態・動態）調査
提供分類2	閲覧
提供周期	年次

表番号	統計表	調査年月	公開（更新）日	表示・ダウンロード
E1	病院数；病床数，病院－病床の種類・二次医療圏・市区町村別	2017年	2018-12-27	CSV DB
E2	一般診療所数；歯科診療所数；病床数，病床の有無・二次医療圏・市区町村別	2017年	2018-12-27	CSV DB
E3	病院数；病床数，病棟等の種類・二次医療圏別	2017年	2018-12-27	CSV DB
E4	病院数（重複計上），診療科目・二次医療圏別	2017年	2018-12-27	CSV DB

平成29年医療施設（静態・動態）調査＞閲覧
E3「病院数；病床数，病棟等の種類・二次医療圏別」

基本マスタ：二次医療圏マスタを作ろう

	A	B	C	D	E	F	
1	平成29年	医療施設調査	平成29（2017）年10月1日				
2	閲覧　第　3表　病院数；病床数，病棟等の種類・二次医療圏別						
3		総数		回復期リハビリ		回復期リハビリ	
4				テーション病棟		テーション病棟	
5				（一般病床）		（療養病床）	
6				を有する病院		を有する病院	
7		施設数	病床数	施設数	病床数	施設数	病
8							
9	01　北海道	561	94523	21	900	26	
10	0101　南渡島	36	7268	1	50	4	
11	0102　南檜山	5	443	－	－	－	－
12	0103　北渡島桧	7	1044	－	－	－	－
13	0104　札幌	235	41563	13	637	10	
14	0105　後志	22	3974	2	79	1	
15	0106　南空知	19	2879	1	24	－	－
			3117	－	－	1	

使うのはここだけ！

基本マスタ：二次医療圏マスタを作ろう

	A	B	C
1	二次医療圏コード	二次医療圏	
2	0101	南渡島	
3	0102	南檜山	
4	0103	北渡島檜山	
5	0104	札幌	
6	0105	後志	
7	0106	南空知	
8	0107	中空知	
9	0108	北空知	
10	0109	西胆振	

4桁の地域コード

完成予想図

基本マスタ：二次医療圏マスタを作ろう

	A	B
1	01　北海道	
2	0101　南渡島	
3	0102　南檜山	
4	0103　北渡島檜山	
5	0104　札幌	
6	0105　後志	

コードと名称を分離させるには？

基本マスタ：二次医療圏マスタを作ろう

1

メニューから

データ > 区切り位置

・「区切り記号付き」を選択

2

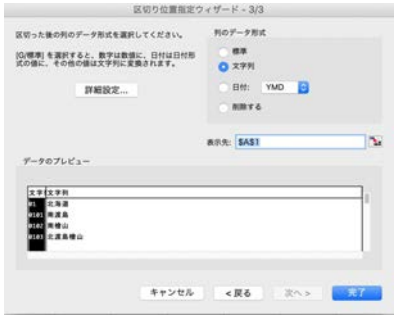
区切り文字：**スペース**

※状況に合わせましょう

3

列のデータ形式：**文字列**

※分割後のコードが入る列は特に重要です



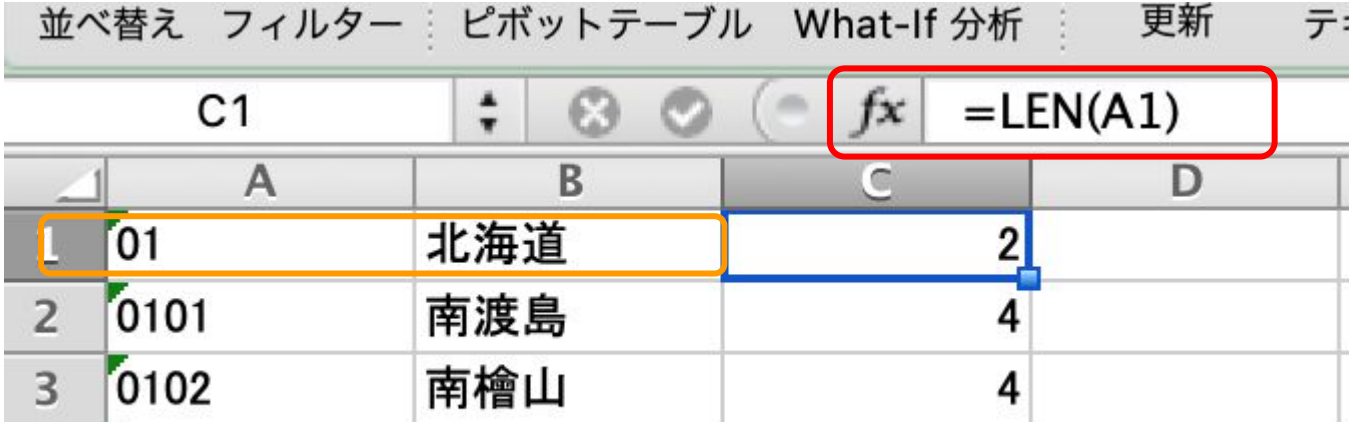
基本マスタ：二次医療圏マスタを作ろう

	A	B	C
1	01	北海道	
2	0101	南渡島	
3	0102	南檜山	
4	0103	北渡島檜山	
5	0104	札幌	

シート名：MASTER

コードと名称が分離できました

基本マスタ:二次医療圏マスタを作ろう



不要な行(都道府県)を削除するには？

基本マスタ：二次医療圏マスタを作ろう

	A	B	C
1	二次医療圏コード	二次医療圏	
2	0101	南渡島	
3	0102	南檜山	
4	0103	北渡島檜山	
5	0104	札幌	
6	0105	後志	
7	0106	南空知	
8	0107	中空知	

見出しをつけて完成です

オープンデータの下処理（初級編）：横断データ1（ SCR）

SCRデーター一覧

- SCR活用の留意点等（平成29年度、平成28年度診療分共通）　(PDF形式：823KB)　※データをご利用になる前

平成29(2017)年度診療分

都道府県	二次医療圏	
診療行為区分　(CSV形式：541KB) 診療行為区分　(Excel形式：540KB)	診療行為区分　(CSV形式：2,594KB) 診療行為区分　(Excel形式：3,434KB)	診療行為区分 診療行為区分
診療行為枝番　(CSV形式：693KB)	診療行為枝番　(CSV形式：3,092KB)	診療行為枝番

内閣府 経済財政諮問会議「医療提供状況の地域差」
https://www5.cao.go.jp/keizai-shimon/kaigi/special/reform/mieruka/tiikisa.html

オープンデータの下処理（初級編）：横断データ1（ SCR）

H29_二次医療圏別SCR_診療行為区分

コード表用番	区分番号	名称	入外区分	0101 南渡島	0102 南檜山	0103 北渡島檜山	0104 札幌	0105 後志	0106 南空知
A	000	初診	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
A	000	初診	2	92.1	42.3	47.2	96.2	73.5	76.8
A	000	初診	3	93.0	42.0	47.7	96.8	73.8	76.5
A	001	再診	2	96.9	50.9	35.4	85.8	87.6	82.1
A	002	外来診療料	1						
A	002	外来診療料	2	101.8		145.4	102.7	80.5	82.2
A	002	外来診療料	3	101.8		145.4	102.7	80.5	82.2
A	100	一般病棟入院基本料	1	151.0	107.7	169.8	144.5	98.9	120.9
A	101	療養病棟入院基本料	1	73.2	110.4	234.9	161.5	117.8	103.6

縦並びにするには？

オープンデータの下処理（初級編）：横断データ1（ SCR）

1

・不要な部分（表の名前など）はカット

・必要な情報だけ事前に抽出

※今回は「コード表 =E」、「入外区分 =1」に絞りました

	A	B	C	D	E	
					0101	0102
2	コード表用	区分番号	名称	入外区分	南渡島	南檜山
2	E	000	遠隔画像診断管理加算２	1	131.5	
5	E	001	他医撮影の写真診断（乳房撮影）	1	143.6	
8	E	002	心臓及び冠動脈造影（左心）（新生	1	146.2	
1	E	003	造影剤注入（静脈造影カテーテル法	1	161.6	
4	E	004	基本的エックス線診断料（４週間以	1	13.8	
7	E	100	シンチグラム（全身）	1	261.8	

2

形式を選択してペースト

「行列を入れ替える」

オープンデータの下処理（初級編）：横断データ1（ SCR）

	A	B	C	D
1		コード表用番号	E	E
2		区分番号	000	001
3		名称	遠隔画像診断管理加	他医撮影の写
4		入外区分	1	
5	0101	南渡島	131.5	
6	0102	南檜山	43.4	
7	0103	北渡島檜山	82.8	

シート名：**SCR**

縦横が入れ替わりました

マスタとの連結：横断データ1（SCR）

基本マスタ

	A	B
1	二次医療圏コード	二次医療圏
2	0101	南渡島
3	0102	南檜山
4	0103	北渡島檜山
5	0104	札幌
6	0105	後志
7	0106	南空知
8	0107	中空知
9	0108	北空知

連結したい項目見出しを追加

	A	B	C	D
			SCR（診療行為区分・入院）H29	
	二次医療圏コー	二次医療圏	遠隔画像診断	他医撮影の写	心臓
	0101	南渡島		
	0102	南檜山		
	0103	北渡島檜山		
	0104	札幌		

マスタとの連結：横断データ1（SCR）

C3 =VLOOKUP(A3,SCR!A5:C348,3,FALSE)

	A	B	C	D	E	F
1			SCR（診療行為区分・入院）H29			
2	二次医療圏コー	二次医療圏	遠隔画像診断	他医撮影の写	心臓及び冠動	造影剤注入
3	0101	南渡島	131.5			
4	0102	南檜山				

【本日の難所】

MASTERシートにSCRシートからデータをもってくる！

マスタとの連結：横断データ1（SCR）

=VLOOKUP(対象文字,検索範囲,出力列位置,FALSE)

MASTERシート**A3**セルの「**0101**」

おまじない！

	B	C	D
	-ド表用番号	E	E
2	区分番号	000	001
3	名称	遠隔画像診断管理加	他医撮影の写真診断
4	入外区分	1	1
5	0101	南渡島	131.5	143.6
6	0102	南檜山	43.4	57.5
7	0103	北渡島檜山	82.8	112.3

SCRシートの範囲**A5:C348**

範囲内の左から**3番目**の列

マスタとの連結：横断データ1（SCR）

	A	B	C	D	E	F	G
1			SCR（診療行為区分・入院）H29				
2	二次医療圏コード	二次医療圏	遠隔画像診断	他医撮影の写	心臓及び冠動	造影剤注入（	基本的エッ
3	0101	南渡島	131.5	143.6	146.2	161.6	1
4	0102	南檜山	43.4	57.5	57.1	9.4	
5	0103	北渡島檜山	82.8	112.3	111.7	31.5	
6	0104	札幌	138.8	153.3	156.1	175	1
7	0105	後志	82.7	95.7	94.9	64.6	

MASTERシートにSCRデータをインポートしました

オープンデータの下処理（中級編）：横断データ2（がん検診）

データセット一覧

<戻る

一覧形式で表示

政府統計名	地域保健・健康増進事業報告
提供統計名	地域保健・健康増進事業報告（地域保健・老人保健事業報告）
提供分類1	平成29年度地域保健・健康増進事業報告
提供分類2	閲覧（健康増進編）市区町村表
提供周期	年度次

25-1	平成２８年度における大腸がん検診受診者数・要精密検査者数・精密検査受診の有無別人数・偶発症の有無別人数，市区町村、検診方式・検診回数・年齢階級別（男総数）	2017年度	2019-03-13	CSV

平成29年度地域保健・健康増進事業報告＞閲覧（健康増進編）市区町村表
25-1「平成28年度における大腸がん検診受診者数・要精密検査者数・精密検査受診の有無別人数・偶発症の有無別人数，市区町村、検診方式・検診回数・年齢階級別（男総数）」

オープンデータの下処理（中級編）：横断データ2（がん検診）

	注：1）精密検査受診の有無別人数については、計数不詳の市区町村があ				
		受診者数			
		総数	40～44歳	45～49歳	50～54歳
	00　全　国	3349408	152020	144634	14248
	01　北海道	96369	4623	4507	460
0	01100札幌市	24704	893	987	85
1	01202函館市	2113	123	95	10
2	01203小樽市	1387	46	48	6

単位を二次医療圏にするには？

オープンデータの下処理（中級編）：横断データ2（がん検診）

・市区町村データ以外は不要
　→「全国」、都道府県の列は削除

・市区町村コード（5桁）の列と
　市区町村名列に分離

※本工程の解説は省略

		総数	40~
01100	札幌市	24704	
01202	函館市	2113	
01203	小樽市	1387	
01204	旭川市	5086	
01205	室蘭市	798	
01206	釧路市	2231	
01207	帯広市	4388	
01208	北見市	2011	

シート名：**WRK**

オープンデータの下処理（中級編）：横断データ2（がん検診）

市区町村コードと市区町村名

01100	札幌市
01202	函館市
01203	小樽市
01204	旭川市
01205	室蘭市
01206	釧路市

基本マスタ（二次医療圏）

	A	B	C
1	二次医療圏コード	二次医療圏	
2	0101	南渡島	
3	0102	南檜山	
4	0103	北渡島檜山	
5	0104	札幌	
6	0105	後志	
7	0106	南空知	

結びつけるには？

オープンデータの下処理（中級編）：市区町村 －二次医療圏対応表

	A	B	C	D
1	市区町村コード	市区町村	二次医療圏コード	二次医療圏
2	01202	函館市	0101	南渡島
3	01236	北斗市	0101	南渡島
4	01331	松前町	0101	南渡島
5	01332	福島町	0101	南渡島
6	01333	知内町	0101	南渡島
7	01334	木古内町	0101	南渡島
8	01337	七飯町	0101	南渡島
9	01343	鹿部町	0101	南渡島

完成予想図

市区町村−二次医療圏対応表をつくろう

オープンデータの下処理（中級編）：市区町村 －二次医療圏対応表

データセット一覧

戻る

政府統計名	医療施設調査
提供統計名	医療施設調査
提供分類1	平成２９年医療施設（静態・動態）調査
提供分類2	閲覧
提供周期	年次

表番号	統計表	調査年月	公開（更新）日	表示・ダウンロード
E1	病院数；病床数，病院－病床の種類・二次医療圏・市区町村別	2017年	2018-12-27	CSV DB
E2	一般診療所数；歯科診療所数；病床数，病床の有無・二次医療圏・市区町村別	2017年	2018-12-27	CSV DB
E3	病院数；病床数，病棟等の種類・二次医療圏別	2017年	2018-12-27	CSV DB
E4	病院数（重複計上），診療科目・二次医療圏別	2017年	2018-12-27	CSV DB

平成29年医療施設（静態・動態）調査＞閲覧
E1「病院数；病床数，病院－病床の種類・二次医療圏・市区町村別」

オープンデータの下処理（中級編）：市区町村 －二次医療圏対応表

1

必要な箇所だけ抽出

閲覧　第　　1表　病院数；病床数，病院			
	病　院		
	施設数		精神科病　院
01　北海道		561	
0101　南渡島		36	
01202 函館市		29	
01236 北斗市	－		－
01331 松前町		1	－
01332 福島町	－		－
01333 知内町	－		－
01334 木古内町		1	－

2

コードと名称を分離

	A	B
1	01	北海道
2	0101	南渡島
3	01202	函館市
4	01236	北斗市
5	01331	松前町
6	01332	福島町
7	01333	知内町

→都道府県、二次医療圏、市区町村が混在

オープンデータの下処理（中級編）：市区町村 －二次医療圏対応表

1

二次医療圏コードと二次医療圏のみ右列に抽出

	01	北海道	0	0
	0101	南渡島	0101	南渡島
	01202	函館市	0	0
	01236	北斗市	0	0
	01		0	
	01		0	
	01		0	
	01334	木古内町	0	0
	01337	七飯町	0	0
0	01343	鹿部町	0	0
1	01345	森町	0	0
	0102	南檜山	0102	南檜山

=IF(len(A2)=4,A2,)

=IF(len(A2)=4,B2,)

オープンデータの下処理（中級編）：市区町村 －二次医療圏対応表

2

メニュー > 選択と検索

>「**条件を選択してジャンプ**」

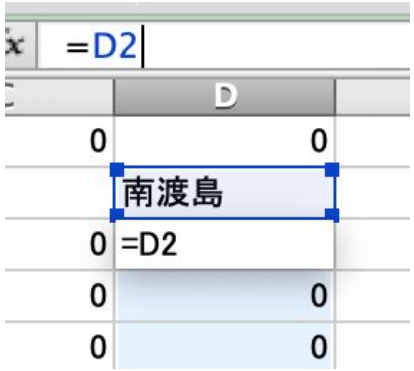
> 数式 > **数値**を選択

3

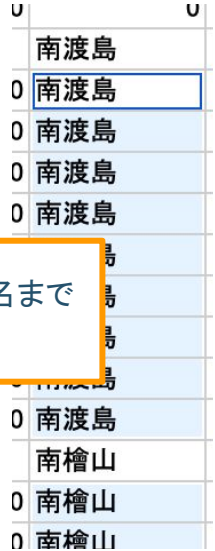
選択されたセルに対して、

「一つ上のセル名」を入力

→ [Ctrl]+[Enter]で確定



次の二次医療圏名までのスキマが埋まる



オープンデータの下処理（中級編）：市区町村 －二次医療圏対応表

	A	B	C	D
1	市区町村コード	市区町村	二次医療圏コード	二次医療圏
2	01202	函館市	0101	南渡島
3	01236	北斗市	0101	南渡島
4	01331	松前町	0101	南渡島
5	01332	福島町	0101	南渡島
6	01333	知内町	0101	南渡島
7	01334	木古内町	0101	南渡島
8	01337	七飯町	0101	南渡島
9	01343	鹿部町	0101	南渡島
10	01345	森町	0101	南渡島
11	01361	江差町	0102	南檜山

シート名：**SIKU**

市区町村−二次医療圏対応表の完成

オープンデータの下処理（中級編）：横断データ2（がん検診）

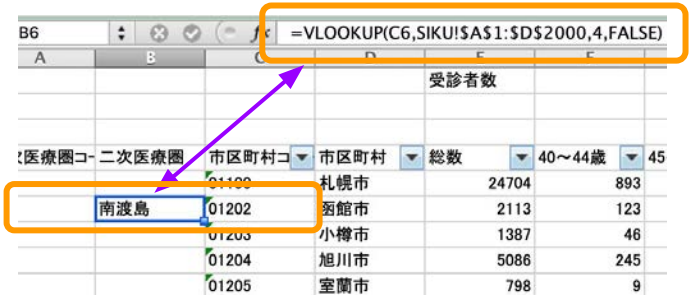
1

WRKシートに見出しを追加

	A	B	C	D	E
					受診者数
	二次医療圏コード	二次医療圏	市区町村コード	市区町村	総数
			01100	札幌市	24704
			01202	函館市	2113
			01203	小樽市	1387
			01204	旭川市	5086
			01205	室蘭市	798

2

VLOOKUPで二次医療圏情報をインポート



オープンデータの下処理（中級編）：横断データ2（がん検診）

	二次医療圏	二次医療圏	市区町村コ	市区町村
5	#N/A	#N/A	01100	札幌市
6	0101	南渡島	01202	函館市
7	0105	後志	01203	小樽市
8	0112	上川中部	01204	旭川市
9	0109	西胆振	01205	室蘭市

インポートできない？

オープンデータの下処理（中級編）：横断データ2（がん検診）

二次医療圏コー	二次医療	市区町村コー	市区町村
#N/A	#N/A	01100	札幌市
#N/A	#N/A	04100	仙台市
#N/A	#N/A	11100	さいたま市
#N/A	#N/A	12100	千葉市
#N/A	#N/A	14100	横浜市
#N/A	#N/A	14130	川崎市
#N/A	#N/A	14150	相模原市
#N/A	#N/A	15100	新潟市
#N/A	#N/A	22100	静岡市
#N/A	#N/A	22130	浜松市
#N/A	#N/A	23100	名古屋市
#N/A	#N/A	26100	京都市
#N/A	#N/A	27100	大阪市
#N/A	#N/A	27140	堺市
#N/A	#N/A	28100	神戸市
#N/A	#N/A	33100	岡山市
#N/A	#N/A	34100	広島市
#N/A	#N/A	39878	中芸広域連合
#N/A	#N/A	40100	北九州市
#N/A	#N/A	40130	福岡市
#N/A	#N/A	43100	熊本市

1

政令指定都市：対応表では区ごとで記載

→対応する医療圏が一意なら手動で入力

2

広域連合：行政統計には稀にみられる

→対応する医療圏が一意なら手動で入力

3

その他の場合

→空欄に

オープンデータの下処理（中級編）：横断データ2（がん検診）

二次医療圏	二次医療圏	市区町村コ	市区町村
0104	札幌	01100	札幌市
0403	仙台	04100	仙台市
1104	さいたま	11100	さいたま市
1201	千葉	12100	千葉市
		14100	横浜市
		14130	川崎市
1410	相模原	14150	相模原市
1502	新潟	15100	新潟市
2205	静岡	22100	静岡市
2208	西部	22130	浜松市
2301	名古屋	23100	名古屋市
2604	京都・乙訓	26100	京都市
2708	大阪市	27100	大阪市
2706	堺市	27140	堺市
2801	神戸	28100	神戸市
3301	県南東部	33100	岡山市
3401	広島	34100	広島市
3901	安芸	39878	中芸広域連合
4012	北九州	40100	北九州市
4001	福岡・糸島	40130	福岡市
4301	熊本	43100	熊本市

「横浜市」、「川崎市」は構成区が異なる医療圏に属するため、一意な医療圏が定まらない

→「市」の単位が「医療圏」より大きい例外

※2018年4月から、横浜市内の三つの医療圏は統合されて一つの「横浜」医療圏になりました

横断データ集を完成させよう：マスタとの連結

	A	B	C	D	E	
1	二次医療圏コード	二次医療圏	市区町村コード	市区町村	受診者数_総数	精密検査受診の有
2	0104	札幌	01100	札幌市	24704	
3	0101	南渡島	01202	函館市	2113	
4	0105	後志	01203	小樽市	1387	
5	0112	上川中部	01204	旭川市	5086	
6	0109	西胆振	01205	室蘭市	798	
		釧路	01206	釧路市	2231	
		十勝	01207	帯広市	4388	
9	0117	北網	01208	北見市	2011	
10	0106	南空知	01209	夕張市	209	
11	0106	南空知	01210	岩見沢市	1757	
12	0117	北網	01211	網走市	793	
13	0115	留萌	01212	留萌市	301	
14	0110	東胆振	01213	苫小牧市	2228	

同じ医療圏が複数

シート名：**WRK**

基本マスタと連結するには？

横断データ集を完成させよう：マスタとの連結

MASTERシート

		SCR（診療行為区分・入院	
二次医療圏コー	二次医療圏	遠隔画像診断	他医撮影
0101	南渡島	131.5	1
0102	南檜山	43.4	
0103	北渡島檜山	82.8	1
0104	札幌	138.8	1
0105	後志	82.7	
0106	南空知	76.9	
0107	中空知	141.8	1
0108	北空知	93.1	
0109	西胆振	113	1

＋

WRKシート

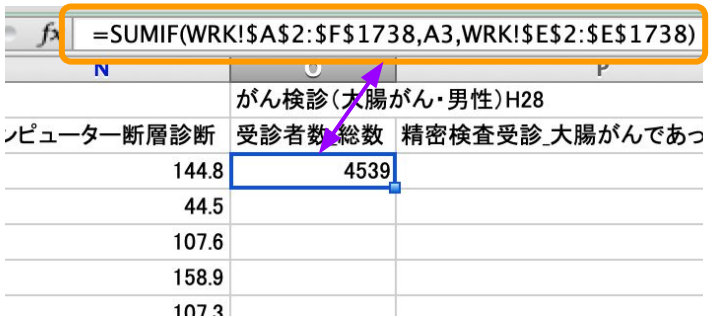
A	B	C	
二次医療圏コード	二次医療圏	市区町村コード	市区
0104	札幌	01100	札幌
0101	南渡島	01202	函館
0105	後志	01203	小樽
0112	上川中部	01204	旭川
0109	西胆振	01205	室蘭
0120	釧路	01206	釧路
0119	十勝	01207	帯広
0117	北網	01208	北見
0106	南空知	01209	夕張
0106	南空知	01210	岩見
0117	北網	01211	網走

横断データ集を完成させよう：マスタとの連結

MASTERシートの右側にWRK用の項目見出しを追加

	A	B	N	O	P
1				がん検診（大腸がん・男性）H28	
2	二次医療圏コー	二次医療圏	コンピューター断層診断	受診者数_総数	精密検査受診_大腸がん
3	0101	南渡島	144.8		
4	0102	南檜山	44.5		
5	0103	北渡島檜山	107.6		
6	0104	札幌	158.9		
7	0105	後志	107.3		
8	0106	南空知	95.7		

SUMIF関数でWRKシートからデータを集計インポートする



横断データ集を完成させよう：マスタとの連結

=SUMIF(検索範囲,抽出対象,合計範囲)

WRKシートの対象範囲

二次医療圏コード

	二次医療圏			受診者数_総数
	0104	札幌		24704
	0101	南渡島	01202	函館市	2113
	0105	後志	01203	小樽市	1387
	0112	上川中部	01204	旭川市	5086
	0109	西胆振	01205	室蘭市	798
	0120	釧路	01206	釧路市	2231
	0119	十勝	01207	帯広市	4388
	0117	北網	01208	北見市	2011
	0106	南空知	01209	夕張市	209
	0106	南空知	01210	岩見沢市	1757
2	0117	北網	01211	網走市	79

集計したい値のある範囲

横断データ集を完成させよう：マスタとの連結

A	B	N	O	P
			がん検診（大腸がん・男性）H28	
二次医療圏コー	二次医療圏	コンピューター断層診断	受診者数_総数	精密検査受診_大腸がんであった者_総数
0101	南渡島	144.8	4539	18
0102	南檜山	44.5	654	1
0103	北渡島檜山	107.6	1111	5
0104	札幌	158.9	31720	89
0105	後志	107.3	4333	19

二次医療圏単位の集計結果がインポートできました

横断データ集を完成させよう：マスタとの連結【注意点】

17	1312	北多摩北部	85.8	22071	95
18	1313	島しょ	39.8	842	1
19	1401	横浜北部	67.9	0	0
20	1402	横浜西部	89.3	0	0
21	1403	横浜南部	87.2	0	0
22	1404	川崎北部	81.4	0	0
23	1405	川崎南部	124.3	0	0
24	1406	横須賀・三浦	75.6	21366	90

「横浜市」、「川崎市」の区部からなる医療圏は、WRKシートに対応する値がないため、SUMIF関数は0を返す

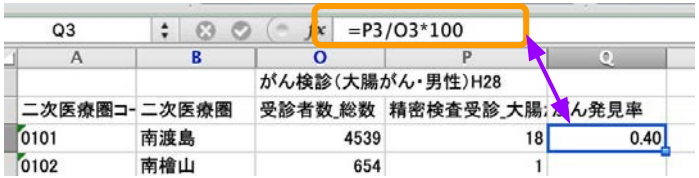
→ 手動で空欄にする

横断データ集を完成させよう

がん発見率（アウトカム指標）
=がんであった者の数 / 検診受診者数×100

「割合」など一部データには、市区町村単位の値を合計して二次医療圏の値とすることはできないものがある

必要な要素を集計後にあらためて計算



横断データ集を完成させよう

	A	B	C	D	E	M	N	O	P
1			SCR(診療行為区分・入院)H29					がん検診(大腸がん・男性)H28	
2	二次医療圏コー	二次医療圏	遠隔画像診断	他医撮影の写	心臓及び冠動	MRI撮影(3テ	コンピューター	がん発見率(%)	
3	0101	南渡島	131.5	143.6	146.2	181.1	144.8	0.40	
4	0102	南檜山	43.4	57.5	57.1	25.7	44.5	0.15	
5	0103	北渡島檜山	82.8	112.3	111.7	66	107.6	0.45	
6	0104	札幌	138.8	153.3	156.1	191.6	158.9	0.28	
7	0105	後志	82.7	95.7	94.9	119.8	107.3	0.44	
8	0106	南空知	76.9	92.8	94.1	113.5	95.7	0.12	
9	0107	中空知	141.8	119.5	121.1	130.7	123	0.58	
10	0108	北空知	93.1	90.6	89.1	56.2	116.9	0.35	

二次医療圏別の横断データ集の完成