

記者ゼミIT講座2022

「見つけた」を「使える」にしよう

このデータを使って、何か発見がないか、いろいろ試してみたい。そんな気持ちになれる宝の山に出くわしたとしましょう。ですが、そのまま使えるとは限りません。たとえば、

- ・うまく読み込めない。必要な部分を取り出せない
- ・複数のソースから集めてきたデータを自分でマッチングさせないとダメ

といった具合で……。

おなじみの表計算ソフトは、こういう「データ整形」まわりのことをやらせるのに、役に立ちます。一緒に試してみましょう。

講師はExcelを使います。ただし、普段からGoogleスプレッドシートをお使いの方や、MacでNumbersをお使いの方は、わざわざExcelをインストールしていただく必要はありません。また、Excelのバージョンも問いません。その代わり「やってみたら、その通りにならないんだけど」というフィードバックを積極的にお願いします。

たぶん、散らかった台所をそのままお目にかかる感じになると思います。正直、とても恥ずかしいのですが、データの料理の仕方・考え方をお伝えするには、それが一番よいからです。

会場に来ていただける方もオンライン参加の方も、困ったら、ストップをかけて下さい。オンラインでは反応が分からないので、講師も不安です。助けるつもりで、頼まれて下さい。逆に、こうしたらうまく行った、というアドバイスやお裾分けも、遠慮せずをお願いします。

資料の置き場所

資料一式はこちらにあります。

<https://github.com/nishioWU/JNPC2302>

右上の緑色の「<> Code」のボタンから、「Download ZIP」を選んで下さい。ダウンロードしたら、必ず解凍を。解凍しなくても中身が見えることがありますが、その後がうまくいきません。解凍したフォルダ「JNPC2302-main」を丸ごと、デスクトップなどアクセスしやすい場所にコピーしておいて下さい。失敗しても、元のファイルをまたコピーしてくれば、やり直せます。

きょうのメニュー

0. Excelの超基本

1. 絶対参照やショートカットキーの練習

「1)表の練習.xlsx」を使います。余裕があったら、事前にトライしてみてください。

2. PDFファイルの読み込みとマッチングの練習

「2)国連決議」のフォルダを使います。練習用のExcelファイルのほか、データのフォルダがあります。フォルダ「UN」の中に、読み込むPDFが入れてあります。

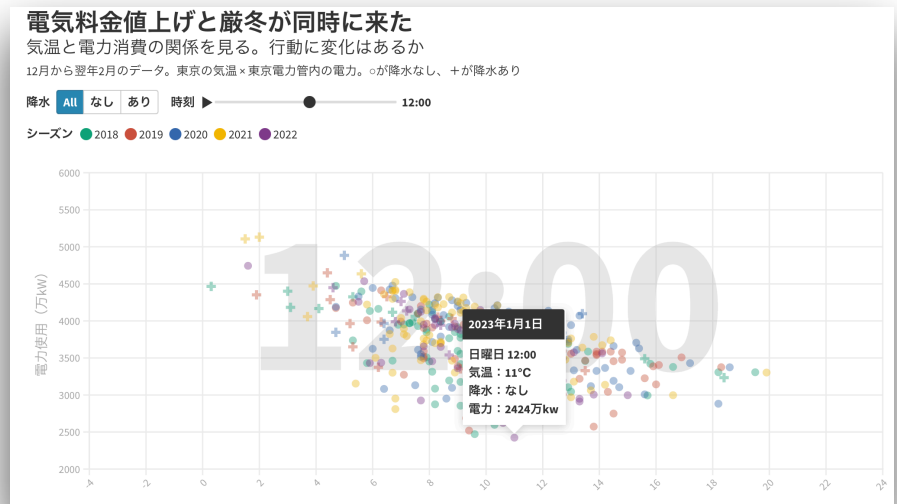
3. 気温と電力の関係をみてる

これがメインです。こちらのビジュアライゼーションの元データを作ります。複数のCSVファイルを読み込むところから始めます。

bit.ly/FUYU-NO-DENRYOKU

可視化に使ったのは、記者ゼミでもたびたび紹介されているFlourishです。今回はFlourishの説明はしませんが、散布図の雛形を選び、CSVデータをアップロードすることで、作成可能です（多少の設定は必要です）。

4. 文字情報がなく、画像になってしまっているファイルだったら時間がなければ割愛します。



0. Excelの超基本

▶行と列

横を行、縦を列とかカラムと呼んで区別している。

どっちが行で、どっちが列か、こんな覚え方も。数学の行列（マトリックス）と同じ。

▶式は小文字で入力しよう

Excelは（Rやほかのプログラム言語と違って）大文字でも小文字でも命令を聞いてくれる。なので、セルに入れて計算を指示する式は、小文字で入力するとよい。正しく認識されれば大文字に変換される。小文字のままで残ったら、打ち間違いがある、ということだから見直そう。

関数入力を手伝ってくれるツール（「fx」ボタン）を使うのもよい。

▶絶対参照

式をコピーすると、気を利かして、計算対象の行や列をずらしてくれる。それが便利だからだが、困る場合もある。そのときは、ずらされては困るものに「\$」マークをつけると、コピー先でもずれない。これが絶対参照。式の入力中に「F4」キーを押すと、行と列の両方またはどちらかに、\$がついたり消えたりして切り替えられる。自動でスライドする、デフォルトの挙動は「相対参照」という。

▶名前

特定のセルや範囲や数式に、分かりやすい名前を自分でつけてやることができる。この名前を使って指定すれば、相対参照でずれることはない（絶対参照になる）。

行列

1. 絶対参照とショートカットキー

Excelを使いこなしている方には常識。が、知らずに長い間使っている方も結構いる。なので、これだけでも覚えて。

①なるべく楽しく作るには

「1)表の練習.xlsx」のファイルを開く。1枚目のシートの表を、楽しく完成させるにはどうしたらよいか。

使うのは「&」。以下のようにすると、最初のセルの内容と、次のセルの内容をつなげたものが、この数式を入れたセルに表示される。このように、Excelに仕事をさせる関数（数式）を使うときは、先頭にイコールをつける。セル名だけでなく、二重引用符で囲んだ文字列も使える。3つ以上つなげることもできる。

=セル名や文字列&セル名や文字列……

または、同じように働く、CONCAT関数を使ってもよい。

=CONCAT(セル名や文字列, セル名や文字列……)

②シートの2枚目の式を増殖させる

式を入れるのは1か所、E7だけにする。このセルに入れておいた式は

E7
=E6&D7

なのだが、これを増殖させてみると、どうなるだろう。右下隅をつかんで右や下にドラッグすれば数式がコピーされていく。セルの右下にカーソルを合わせてダブルクリックして、下方向に一気にコピーするやり方もある。すると……想定外の結果になってしまう。

それぞれのセルに入った数式を見てみよう。Excelが気を利かせて、参照元のセルをスライドしてくれているせいだ、ということが分かる。これは相対参照といい、本来は優れたものの機能だが、この場合は困る。

③絶対参照

そこで重宝するのが、絶対参照。ファンクションキーの「**F4**」を使う。数式の中のセル名の部分にカーソルがあると、F4キーを押すたびに、**\$マーク**がついたり消えたりする。\$がついたところは、ほかのセルに数式を増殖させたときにも、スライドされずに固定になる。元の式は同じE6であっても、\$マークの付け方を変えると

\$E\$6 どのセルにコピーしても、必ずE6を指し示す

\$E6 縦の列はEのままで固定だが、横の行はスライドされる

E\$6 縦の列はスライドされるが、横の行は6行目のままで固定

という指示になる。

では、E7セルに入っている数式を、「**F4**」キーで修正しよう。数式の中の「E6」の部分にカーソルが来ている状態でF4キーを何度か押し、「D7」にカーソルが来ている状態で同様にF4キーを何度か押して

E7

=E\$6&\$D7

とする。これで、青と赤の網掛け部分を正しく指定するようになった。この数式を、右と下に増殖させて、確かめてみよう。④のシートのようになれば、正解。

④シートの複写

シートを丸ごと複写して増やす練習をする。シート名が表示されているタブ（耳みたいな部分）を、**Ctrlキーを押しながらドラッグ**する。そうすると、まったく同じシートがもう1枚できる。

⑤「値だけ」の貼り付け

増殖させたほうのシートを選ぶ。数式を消して、ただの文字列に置き換えたい。つまり、Excelに仕事をさせた後の結果だけを保存したい。

表の中のどのセルでもよいので指して、そこで「**Ctrl + A**」または「**Ctrl + ***」。これで、表の範囲だけが指定されたはず。

そうしたら、「**Ctrl + C**」でコピーして、「**Alt + Ctrl + V**」。単独で「**V**」。最後に「**Enter**」キー。これで、数式は消えて、結果の値だけになった。よく使うので、ぜひ覚えて。

⑥復習

⑥のシートで、ここまでの復習を。O列の上あたりに見えるプラスマークを押すと、答え合わせができる。

2. ファイルの読み込みとマッチング

▶ファイルの読み込み

Excelのメニュー欄の「データ」>>「データの取得」と進んで、該当するものを選ぶ。データの受け渡しによく使う形式の「CSV」は、目立つところにあるはず。一方、さらに下位の「ファイルから」に進むと、「PDFから」や「フォルダーから」も選べる。

▶VLOOKUP関数を使ったマッチング

=VLOOKUP(セル名, 検索範囲, 列数, FALSE)

セル名 このセルの内容と一致するものを

検索範囲 ここの左端の列をターゲットにして探し出す。勝手にスライドされないよう注意。「数式」>>「名前の管理」>>「数式で使用」にある「名前」で指定することも可能。ショートカットキー「**Ctrl + F3**」で、「名前の管理」が開く。範囲指定した後、**画面左上の「名前ボックス」で命名する**のが簡単。

列数 一致したところを起点に数え始め、右にこの分だけ進んだものを取ってくる

FALSE 一致するものを探す場合は「FALSE」で決め打ち

という文法。重複データがない、つまり上から下まで見渡したときに同じものが出てこない、一意のデータだけがある「キー列」どうしを使って、2つのデータセットを結びつけるときに使う。ただし、式を増殖させる際に、検索範囲のほう勝手にスライドされると、想定通りの動作にならない。絶対参照や、名前による指定を使おう。セル名のほうは逆に、縦方向にスライドしてくれ

ないと困るので、相対参照にしておく（横方向は固定でよいが、ほかの列で式を流用しないなら、気を使わなくてよい）。

範囲の左端がキー列でない場合は、左端に来るように工夫するか、ほかの関数を使う。進化版のXLOOKUPなら、左端でなくてもOKだし、いちいち「FALSE」を指定する必要もない上に、一度に複数列を取ってこられるメリットがある。ただ、Excelのバージョンによっては使えないので、きょうのところは、VLOOKUPを覚えて慣れよう。使える方は、遠慮せずXLOOKUPを。

▶テーブル化

特定の列に注目した絞り込み（フィルター）や並び替え（ソート）などの機能が使えるテーブルを作る。テーブルにしたい部分で「**Ctrl + T**」を押せばよい。先頭の行を項目の見出しとして使う場合は、チェックボックスで指定する。いきなり「**Ctrl + T**」を押してもよいが、思い通りの範囲にならないこともあるので、「**Ctrl + A**」や「**Ctrl + ***」で先に範囲指定しておいてから、が安心だ。

テーブルにすると、ひとつのセルに式を入れるだけで、列全体に当てはめてくれるようになる。手動で式を増殖させる手間が省けて便利だ。列や行を増やせば、テーブルの範囲も広がる。また、下にスクロールしても見出し行が画面からはみ出ないで残る（自分で「ウインドウ枠の固定」を設定しなくてよい）のも、楽でよい。

2.1. こんなシナリオ

国連総会で対ロシア避難決議の投票が行われた。直近3回で各国の態度がどう変化したかを追って、一覧表や地図にしたい。過去2回分の投票結果はすでに、Excelの表にしてある。つい先ほど、3回目の結果がPDFファイルで広報された。この分を、これまでの表に足したい。

2.2. PDFを読み込む

①PDFファイルの中身を確認

まず、ファイルの中身を確認しておく。「2)国連決議」フォルダの下、「UN」フォルダの中にある「A ES-11 L.5.pdf」というファイルを開き、内容を確認する。全体状況と国別の行動と、計2つの表がある。このうち、国別の表のデータを、これまで2回分のデータと結合したい。

②ファイルを読み込む練習

空っぽのExcelファイルを新規作成。「データ」>>「データの取得」>>「ファイルから」>>「PDFから」と進む。

先ほどの「A ES-11 L.5.pdf」というファイルを指定し、「インポート」。バージョンによっては「外部データ」だったり、「フォルダーパス」と表示されることもある。

ほしいのは国別の表。「Table002(Page 1)」のほうなので、これを指定して「読み込み」。しばらく待つと、取り込まれる。完了後、右の方に「クエリと接続」の窓が開いていたら、閉じてしまってもよい。閉じる前に削除してもよい（複雑な処理を繰り返す場合には、残しておくとう便利。今回は不要）。読み込みの練習なので、このExcelファイルは削除してOK。

③これまでの表と同じExcelファイルに読み込む

今度は「2-1)国連決議_練習用.xlsx」のファイルを開く。これまで2回分のデータがあらかじめ表にしてあるほか、国名を3文字の略称で表したISOコード表も入れてある。

先ほどの②と同じ手順で、PDFから国別の表を読み込む。新しいシートに3回目の国別データが読み込まれた。

2.3. テーブルにする

①4段から1段に

4段の段組み状態では扱いにくいので、1段にする。つまり、縦に積み重ねるようにする。

シートを新規作成して、適宜名前をつける。そこにコピー&ペーストで貼り付けてもよいし、スピルという機能（「1)表の練習.xlsx」の7枚目シートに使用例がある）で読み込んだ後に、「値だけ」を貼り直してもよい。

ここまではすんだものが「2-2)国連決議_PDF読み込み完了.xlsx」なので、読み込みがうまく行かない場合は、そのファイルに乗り換えを。

②テーブル化

縦1段にできたところで、範囲指定した後「**Ctrl + T**」でテーブルにする。最初の列は「態度」とか「アクション」、次の列は「国名」とでもしておく。

2.4 マッチング

①ISOコードとの結合

読み込んだら、過去の結果とすんなり結合できるかといえば、そうは問屋がおろさない。加盟国数に変化はないにも関わらず、並び順がそろっていないからだ。実は、国連の発表は表記の揺れも意外と多くて、悩まされる。

そこで、ISOコードを使うことにした。国ごとに固有の略称を決めてあるもので、過去2回分の結果はすでにこのコードと結合してある。なので、今回読み込んだ表もまず、ISOコードに結びつけた上で、過去の結果とマッチングさせるようにする。3回目結果の国名→これまでの結果の国名、とダイレクトに突き合わせるのではなく、3回目結果の国名→ISOコード→これまでの結果のISOコード、とたどって結合させるわけだ。

国名のセルのすぐ右のセルを指定し、「ISO略称」と入れる。すると、自動的にテーブルの列が増えた。ここに、国名と対応する3文字略称を取ってくるようにしたい。どうすればよいだろう。VLOOKUP関数を使って考えてみて。

C2セルに数式を入れるとすると、

C2

=VLOOKUP(B2, ISOコード!\$C\$2:\$D\$250, 2, FALSE)

となる。B2セルに入っているもの（ここではAFGANISTAN）を、「ISOコード」というシートのC2からD250までの範囲の左端（つまりC2からC250まで）で探して、一致するものが見つかったら、そのすぐ右の列に入っているもの（ここではAFG）を取ってくる、という指示になる。B2のほうは相対参照のままなので、すぐ下の行ではB3、その次はB4……という式になる。一方、検索範囲のほうは\$マークをつけて絶対参照にしているので、こちらは固定だ。ここがポイント。ターゲットにする範囲の左端が、この場合のキー列である国名になっていることも、確かめて。

!の前はシート名。同じシートなら不要だが、別のシートのセルや範囲を指定するときには、このように指定する。ただし、セル名やシート名を一から手入力する必要はなく、数式を入れている最中に、マウスで参照先のセルや範囲を指定してやれば、入力される。

なお、「B2」とセル名を打つ代わりに、マウスでクリックして指定した場合は、「[@国名]」のように列の見出しが入るはずで、分かりやすい。これでも問題なく動く。@がついている場合は、対応する行、というような意味合い。つまり、名前で指定して列は固定なのだが、行のほうは相対参照と同じくスライドしていく、という動作になっている。

②名前をつけよう

それから、断然おすすめなのは、ターゲットにする範囲に、自分で名前をつけておくやり方。範囲指定してから左上の名前ボックスで命名するか、「数式」>>「名前の管理」>>「名前の定義」（「Ctrl + F3」で「名前の管理」まで進むのが早い）と進み、好きな名前をつけてやる。くどくて申し訳ないが、こうすると絶対参照になるのでスライドされる心配がない（検索範囲のほうは、スライド厳禁だった。スライドしてよいのは、セル名のほう）し、何をしているか分かりやすいのもメリットだ。

もし、さっきの式に出てきた検索範囲に「ISOコード表」と名前をつけた場合で、なおかつ、B2セルと打つ代わりにマウスで指定したならば、

C2

=VLOOKUP([@国名], ISOコード表, 2, FALSE)

となる。「ISOコード表」のところは、「数式」>>「名前の管理」>>「数式で使用」と進んで選ぶと確実。途中まで打ちかけると候補が表示されるので、TABキーで選ぶこともできる。

この式がExcelに指示していることは、次の通りだ。「国名」列のセルと同じものを、ISOコード表の左端の列で探してほしい。一致するものが見つかったら、その右の列（左端から数えて2列目）のセル内容を取得して表示せよー。さっきの式よりも、ぐっと分かりやすいのでは？

ここまですんだものが、「2-3)国連決議_ISOコード作業中.xlsx」。

もし、XLOOKUPを使える環境ならば、さらにISOコード表もテーブル化して、もっと解説しやすい式にできる。

C2

=XLOOKUP([@国名], ISOコードテーブル[探す列をここに], ISOコードテーブル[値を返す列をここに])

「2-4)国連決議_最終形.xlsx」のファイルで使っているの、後で見ていただきたい。

列名に@がついていると、対応する行のスライド、の意味。@なしだと、その列全体で固定、という動作になる。セルのほうは@がついているのに対し、検索範囲のほうは@がついていない。ただし、どれも手で打つ必要はなく、マウスでセルや列を指定すれば、自動的に名前を入力されるので、難しく考える必要はない。

③エラー箇所を見つける

大体マッチングしたはずだ。でも、エラーが出ているところもある。

エラーにも何種類かあって、入れた数式が間違っていれば、「#NAME?」というエラー表示が出る（「5号室にいる山田さんを探せ」と指示を出したのに、語彙や文法の間違いで、言葉が通じ

ない状態に相当)。「#REF!」ならば、言葉は通じているが、探す場所の指示間違い。指定したシートが削除されてしまっているときなどに出るエラー(5号室と指示されたけど、4号室までで5号室はなかった、というような状態)。今回は「#N/A」が出ている。これは、意味が通じて、場所も分かって、指示通りの場所を探しに行ったのに、見つからなかった、ということだ(言われた通りに5号室を探したけれど、そこに山田さんという人はいなかった、という状態)。

テーブルの先頭行、「ISO略称」の右の下向き▼をクリックする。すると、絞り込みやソートの小窓が開く。「テキストフィルター」の先頭「(すべて選択)」にチェックがついていたら外し、一番下の「#N/A」だけにチェックを入れる。これで、マッチングがうまくいかなかった国だけが、絞り込み表示された。

下向き▼は、フィルター作動中を示す漏斗のマークに変わっているはず。そこをクリックして、「フィルターをクリア」を選べば、また全体が表示されるようになる。

④エラーを修正する

コンゴ民主共和国、ミクロネシア、セントビンセント・グレナディーンのように、PDFを読み込む際にセル内で改行されてしまったものは、それを修正すればマッチするようになる。ラオスのように一部が略称になっていたためにマッチしなかったところもあるので、手作業で修正する(LAOと直接打ちこむ。そのセルの数式が消えるが、構わない)。米国・英国・韓国・北朝鮮なども、表を見比べて、同様に手作業で修正する。コートジボワールがマッチしないのは、国名のアクセントがなかったためなので、やはり手作業で修正する。

2.5. 置き換える

これで、ISOコードとのマッチングは終わった。

今度は、各国の投票態度を処理していく。記号を日本語に置き換えたい。

①空欄を置き換える

元のPDFを見ると分かるが、態度が空欄の国がいくつかある。仮に「ZZZ」に置き換えておこう。

「態度」の列だけを範囲指定し、「**Ctrl + H**」で置換の小窓を開く。「検索する文字列」は空欄のまま、「置換後の文字列」にZZZと入れ、「すべて置換」。

または、以下のやり方でもよい。見出し横の下向き▼を押し、テキストフィルターで、「(空白セル)」だけにチェックを入れて絞り込み表示する。次に、空白のセル全体を「Shift」キーなどで範囲指定してから、「ZZZ」と打つ。一番上のセルに入力されるのが見えるはず。そして、Enterではなく、「**Ctrl + Enter**」を押す。これで、選択したセルに一度にZZZと入力された。最後に、フィルターを解除しておく。

賛成票を投じるY、反対票を投じるN、棄権のA、無投票(ZZZにした)の計4パターンのアクションがあり、棄権と無投票は別物として扱われている、ということだ。

②態度の列を右端にも作る

3字略称をキー列にして、今処理した3回目の結果を、これまでの2回分の表に結合したい。だが、VLOOKUPの仕様で、検索する範囲の左端がターゲットでなくてはならない。ISOコードのほうが左になればうまくいかないわけだ。それで、態度と同じ中身の列を、右端にも作ってやる。

コピー&ペーストや、ドラッグだと、貼り間違えてずれる危険が常にある。なので、数式を使いたい。何も加工せずに持ってくるだけだから、なんのことはない、D2セルに入れるなら

D2
=A2

でOKだ。マウスで指示した場合には、列名に置き換わるはずだが、それで結構。

③Y, N…を「賛成票」「反対票」…に置き換える

さらに、E列には、D列の「Y」「N」といった表記を日本語に置き換えたものを入れておこう。
「対照表」という名前をつけた表を用意しておいたので、これを使って、E2セルに

E2
=VLOOKUP(D2, 対照表, 2, FALSE)

と入れればOK。D1とE1には、自分で分かればよいので、「態度2」とか、「日本語にした」とか、見出しを入れておこう。

2.6. 再びマッチング

①これまでの2回分に結合する

今度は過去2回の賛否の表のシートを選び、今回分の結果を入れる列を作る。B列のISO略称をキーとして、同じものを3回目決議のシートから探し、今回のアクションを取ってくる、というふうにしたい。2つ右の列（つまり起点となるISOコードの列から数えて3列目）を取ってくるのだから、列数のところに入れる数字は3。なので、F2セルに入れる数式は、

F2
=VLOOKUP(B2, '3回目などのシート名をここに'!\$C\$2:\$E\$194, 3, FALSE)

となる。

このように、シート名が数字で始まっているときは、一重引用符で囲んでやらないとエラーする。これはExcelの仕様。「決議3回目」のように数字が途中であれば、一重引用符は不要。ただし、検索範囲に名前を付けてやるほうが、断然よい。Excelのクセの説明のために、あえて変な例を出しただけ。

②名前をつけるやり方

先ほどの3回目決議のシートに戻り、左端が3字略称の列、右端がアクションの日本語表記の列となるように範囲指定する。C1またはC2から、E194までの範囲だ。このエリアを選択した状態で、左上の名前ボックスに文字を入れるか、「**Ctrl + F3**」から名前の新規作成に進んで、分かりやすくつける。「3回目結合用」とでもしたいのだが、先頭が数字だと受け付けてくれない。これも仕様。仕方がないので、記号でも補って「_3回目結合用」などとしよう。これを使えば

F2
=VLOOKUP(B2, _3回目結合用, 3, FALSE)

と、すっきり指示が出せる。後で自分で見直したときにも、意味が分かりやすい。

2.7. 値だけのシートを作る

3回分の結果をひとまとめにしたシートができた。でも、今はまだ数式が入ったままだ。値だけにしたシートを作ろう。

「**Ctrl + ドラッグ**」でシートを複写。シートの全範囲を選択して、値だけ貼り付け。シート名は適宜付け替えておく。このシートだけをCSV形式で書き出すと、表にしたり、地図にしたり、使い出があるだろう。

書き出すには、「名前を付けて保存」を選び、ファイルのタイプでCSVを選択する。おすすめは「UTF-8」とついているほうだ。

2.8. 国連に注文はないか

お疲れさまでした。休憩にしましょう。

それにしても、どうしてこんなに苦労しないといけなかったのか。投票結果の発表がどんな形式だったら、すんなり結合できたのか、を考えてみて。できれば、こうしてほしいんだけど、と国連の広報担当にお願いしよう。

3. 気温と電力の関係をみる

3.0. シナリオとポイント

ウクライナ情勢などの影響もあって電気料金が高騰している。一方、この冬の寒さは例年よりも厳しい。家計を考えて節約する、といった行動の変化はあるのだろうか。近年の12月～2月の東京の気温と、東京電力管内の消費電力のデータを使って、調べてみたい。気温を横軸、電力を縦軸に取って散布図を描こうと思うが、そのためには、1時間ごとのデータを突き合わせて、気温と電力のペアにしなくてはならない。

▶VLOOKUPによるマッチング

一意のキー列になりうるものはあるか。もしなかったら、作るとは可能か。両方のデータセットのキー列は「同じ」か。

▶ファイルの読み込み

先にExcelを開いておき、そこに複数のファイルをまとめて読み込むようにする。CSVファイルをダブルクリックして、Excelが立ち上がって開く、というやり方は避ける。

▶データ型を意識する

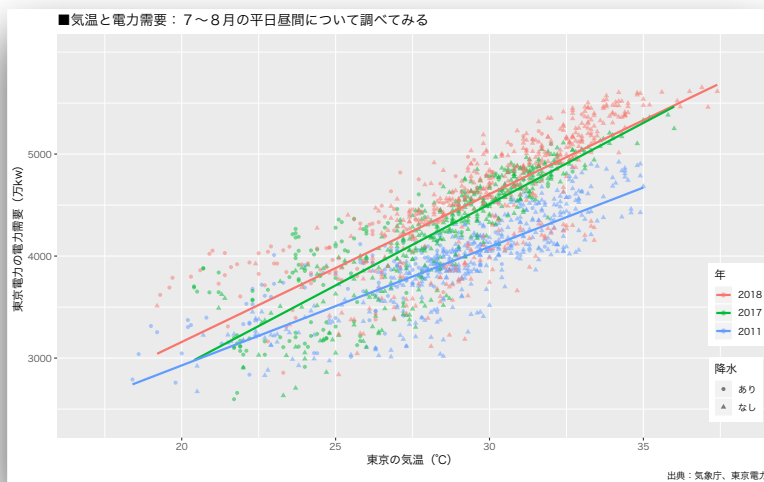
数字の1と文字の"1"は別物。日付と文字列も別物。Excelで不用意に開くと、データ型が変わってしまうことがあるので、要注意。意図してデータ型を変換するための関数もある。

3.1. 夏バージョン

ちなみに、熱中症が命に関わる危険であること、エアコン使用をためらうべきでないこと、の周知が進んだ2018年の夏に、気温と消費電力の関係を調べてみたことがある。平日の日中（リモートワークは今ほど一般的でなかった）のデータに絞ってみると、ほぼ直線的な関係が成り立つことが分かった。しかも、前年とは線の位置や傾きが異なっていた。

また、東日本大震災のあった2011年は、夏にも節電が心がけられていたこと、さらに、そもそも涼しい夏だったらしいことも分かる。

このときの図は「R」で描いたものだが、Excelでも少し工夫すれば同様のものが描ける。回帰直線については、付録を参照のこと。



3.2. データの取得

何が必要かを考える。日時・気温・消費電力のほかは、降水の有無や風速、天気あたりか。

気象庁と東京電力のサイトから、各シーズンのものをダウンロードする。ただし、参加者全員でやるとタイムアウトなどが発生するので、本日は「3)気温と電力」フォルダの下に「JMA」「TEPCO」のフォルダ名で用意しておいた。

どちらも、よく使われるCSV（値をカンマで区切ったもの）形式で、文字コードはShift-JIS。この拡張子はExcelに紐づいていて、ダブルクリックすると立ち上がる設定の方がほとんどだと思うが、ちょっと待った。いきなりExcelで開くと、想定外にデータを変換されてしまうことが多い。郵便番号の先頭の0が取れてしまったり、有効数字を無視して10.0が整数の10になってしまったり、番地が日付に変わってしまったり、背番号の0と00の区別がつかなくなったり、等々。CSV形式のデータは、あらかじめ「メモ帳」などのテキストエディタで開き、ざっと眺めておく習慣をつけるとよい。なお、Excelで開いた場合も、上書き保存せずに、閉じるだけなら壊れないので、焦ることはない。

今回は、気象データのうち気温、降水の有無だけを使うことにした。

こちらは以前、夏のデータでやってみたもの

3.3. CSVデータの一括読み込み

①ファイルの用意

本来であれば、Excelファイルを新規に作り、分かりやすい名前をつけておく。そのファイルを開く。ここに、まず電力、次に気象データを読み込む……のだが、きょうのハンズオンは「3-0) 気温と電力_練習用.xlsx」のファイルを開いて、そこで作業してほしい。説明しやすくするために、他意はない。

ご自身で復習する際には、ぜひ、Excelファイルを新規作成してから、やってみて。

①TEPCOフォルダから読み込む

国連決議のときと途中までは同じで、「データ」>>「データの取得」>>「ファイルから」と進む。ただし、今度はあるフォルダ内の同種のファイルすべて（他の種類のファイルはない想定）を一括して読み込みたいので、「フォルダーから」を選択する。フォルダを選択する小窓が開くので、「3)気温と電力」の下にある「TEPCO」のフォルダを指定して、「開く」をクリックする。

少し時間がかかるかもしれないが、ファイル名の一覧が表示される。「結合」>>「結合および読み込み」を選ぶ。結合の下にあるほうではなくて、最初から見えている「読み込み」を選びがちだが、そうすると、中身ではなくファイル名の一覧が読み込まれてしまうので、注意（そのときは、シートを削除してやり直す）。

サンプルとして、先頭のファイルの中身が表示されるので、「OK」をクリック。少し時間がかかるかもしれない。新しく「TEPCO」というシートが作成され、そこにCSVファイルを全部つなげたものが読みこまれたはずだ。あらかじめテーブル化されている。

②JMAフォルダから読み込む

電力のときと同様に、気象データを読み込む。今度は選ぶフォルダは「JMA」。できあがるシートも「JMA」という名前になる（先に同名のシートがなければ）。やはり、あらかじめテーブル化されている。

データの読み込みがうまくいかなかった方は、「3-1)気温と電力_データ読み込み済み.xlsx」を使って。この後のステップで詰まってやり直す場合にも、このファイルからご一緒に。「3-2)気温と電力_処理の流れと最終形.xlsx」には、中間のステップの区切りのよいところも保存してあるので、途中で追いつきたいときや、処理のフローを見渡したいときには、そちらをどうぞ。

3.4. 「同じ」にするために

では、「3-0)気温と電力_練習用.xlsx」に、データを読み込んだところから。

電力と気象のデータの入っているシートは極力触らないようにして、「スタート!」と名前をつけておいたシートで、マッチングや整形の作業をする。元データを残しておくだけでなく、データが更新された際の手戻りを防ぐ意味もある。

結合する際の手がかりになる、一意のデータが入った列を「キー列」と呼ぼう。気象データのほうには、「年月日時」というキー列があるが、電力にはない。でも、「DATE」と「TIME」をつなげてしまえば、同等のキー列が作れそう。

ところで、一見したところ同じに見えるセルでも、Excel内部のデータ型はいろいろある。型が違えば、同じとは判定されない。数字の1と文字列の1("1")は別物だ。

「日付」形式の場合も、中身は1900年1月1日から数えた日数。つまり数値で、表示するときだけ、日付っぽく見せているだけだ。だから、試しに「数値」形式を指定してやると、正体が見える。

通常の日付なら、日数つまり整数になるので問題はないが、今回のように時刻も関わってくると、要注意だ。1時間単位ならきれいな数字だ、と思うかもしれないが、 $1 \div 24$ は無限小数になる（24は3の倍数）ので、1時間ずつ足していくと誤差が生じる。型も同じなのに、この誤差のために「同じ」「一致した」と判定されない場合がままある。なので、あくまでも見た目通りの「文字列」形式のまま、マッチングさせることにしたい。

①電力データをコピー

「TEPCO」シートのテーブルの、見出し行以外を範囲指定する。つまり、AからD列の、2行目から末尾の行まで（4万4587行）。列はともかく、何万行もスクロールするのは大変だから、ShiftやCtrlキーと矢印キーを上手に使おう。範囲指定したら、「**Ctrl + C**」でコピーする。

②「値のみ」を貼り付け、テーブル化

「スタート!」シートのB10、「ここに貼り付けて」と書いてあるセルに、値のみを貼り付ける。覚えているだろうか。「**Alt + Ctrl + V**」、単独で「**V**」、そして「**Enter**」。B7～E7の色付け

してあるセルをコピーして、B9～E9に貼り、見出しとして使う。さらに、「**Ctrl + T**」でテーブルにしよう。表にする範囲は9行目以下で、9行目を見出しにする。もし、見出しの色が変わってしまったら、もう一回貼り直して、赤くしておいて。

まっさらなシートで始めた場合には、電力のデータを左上隅かその近くに値のみ貼り付けした後で、「**Ctrl + T**」でテーブル化すればよい。

ここまでで、BからEまでの赤い見出しの列に、電力のデータが並んだ。元のまま、とくに加工なしだ。

③不要な行を削る

さて、各時間のデータそのものではなく、項目名だったり注釈だったり、不要な行が少しある。そういう行は、電力使用の数値が入るE列が空欄になっている。そこで、フィルターをかけてE列が空白の行だけに絞りこんで、その行を削除する。

Excelの画面の左端、行番号の色が変わっているのに気付いただろうか。これは、絞り込み表示中であることを示している。テーブルの中のどの列でもよいので、1列だけ選んで、上（見出しは選ばない）から下まで範囲指定する。右クリックから「削除」>>「シート行全体」。何も表示されなくなるが、慌てずにフィルターを解除しておく。

④テーブルにする理由

テーブルにしてしまえば、表の見出し行のすぐ下（行番号だと10行目）に数式を入れた途端、下まで増殖するので、入力が楽だ。うまく入らない場合のみ、セルの右下隅をダブルクリックして、自動入力で下まで増やそう。表のすぐ右側の列に見出しや数式を入れれば、表の範囲が自動的に広がってくれるのも便利。

それから、今やったように、フィルター機能で絞り込みをかけられるのも、強みだ。

⑤電力のキー列を作る

気象データと違って、電力にはキー列がない。同じ「DATE」が24行ずつあるし、同じ「TIME」は無数にある。でも、DATEとTIMEをつなげた列を作れば、同じ日時はひとつもなくて、一意になる。そこで、CONCATを使って、F列をキー列にすることにしよう（&でもよいが、つなげるものが多いときに、何度も&を入力するのが煩わしい）。ただし、単に

F10
=C10&D10

や

=CONCAT(C10, D10)

とするのではダメだ。

日時を表す一意の列にするだけでなく、スタイルも、これからマッチングさせる気象データと同じにしなくてはならない。どうすればよいか、観察すべし。

- ・日付と時刻の間には、半角の空白がある
- ・気象データには、00秒がついている
- ・実はまだ、違う部分があるのだが……

ので、正しくは

F10**=CONCAT(C10, " ", D10, ":00")**

とすることがある。もし、&を使うならば

F10**=(C10&" "&D10&":00")**

となる。

テーブルにしてあるので、数式や見出しを入れた時点で、表が右側に広がった。見出しは引き続き、7行目からコピーして使ってほしい（使わずに、自由に説明を入れても結構）。

3.5. 不要な行の削除

電力のデータは通年なのに対し、気象データは12月から翌年2月まで。なので、3月から11月までの電力データは、気象データのほうにマッチする相手がいない。2018年の1月と2月もそうだ。そこで、先に削ってしまおう。

自分でダウンロードしていれば最初から分かっているが、そうでなければ、この後のマッチングで気づく部分。先取りするのはちょっとずるいのだが、データ量が多いままだとPCがハングするかもしれない。それではハンズオンに困るので、きょうはこの順番にした。悪しからず。

①文字列型から日付型に

冬場以外のデータを捨てる処理をするには、月（今回は年も）を知る必要がある。なので、見た目は日付や時刻のスタイルだが実は単なる文字列、というデータを、日付型に変換してやる。こうすると、年と月を取り出せるようになるので、これをフィルター処理に使うことにする。型の変換には「DATEVALUE」と「TIMEVALUE」、年と月を知るには、それぞれ「YEAR」と「MONTH」を使う。

G10**=DATEVALUE(C10)**

これで、C列では文字列だったものが、G列では日付型になった。もし、数字になってしまう場合は、G列の表示形式を「標準」や「数値」ではなく「短い日付形式」にする。

②年と月を取り出す

H10**=YEAR(G10)****G10****=MONTH(G10)**

これで、年と月を取り出すことができた。もし、^こ五・^{とおび}十日に特有の傾向がないか調べたい、といったニーズがあれば、「DAY」関数で日にちを取り出す列も、同様に作ればよい。

③絞り込んで削除

テーブルのフィルター機能を使う。まず、月の列の見出し、下向き▼のフィルターを開く。数値フィルターで、3から11までにチェックをつけて「OK」を押し、絞り込み表示にする（「指定の範囲内」で、3以上11以下、という指示の仕方も可能。お好みで）。画面左端の行番号の色が変わった。フィルターが効いている証拠だ。

テーブル中のどの列でもよいので1列選んで、先ほどと同じように上から下まで（ただし見出しは除く）選択。右クリックして「削除」>>「シート行全体」。

次に、年と月の列のフィルターを掛け合わせて、2018年かつ1月と2月、の列を同様に削除する。年の列と月の列にそれぞれフィルターをかける。それとも、日付型にしたG列でフィルターをかける（プラスマークを押すと、下の階層が展開される）か、これもお好みで。終わったら、フィルターを解除しておく。

以上、フィルターの練習。この後で「シーズン」の列を作るが、そこで「2017」を選択して削る手もある。

④最後のセルにジャンプ

不要な行を削除したので、Excelのシートが無用に重たくならずにすむはず。でも、もし挙動が変なときには、データの入った最後のセルにジャンプする「**Ctrl + G**」「**Alt + S**」、単独で「**S**」を試すとよい。変なところに飛ぶようなら、何かゴミが残ってしまっているのだから、削除して保存し直すか、最初からやり直すようにする。

3.6. シーズンの列を作る

①条件分岐

年ごとに傾向を見たいのだから、冬という季節は、年をまたいでいる。2022年の12月は、その年の1～2月と同じグループにするよりも、2023年の1月とまとめるべきだろう。そこで、シーズンという列を作ろう。

月の列に「12」の数字が入っていれば、シーズンは「年」そのまま。でも、それ以外なら、「年」から1を引く算した数字にしてやる。こういう、枝分かれがある処理には「IF」を使う。

▶IFによる枝分かれ処理

=IF(分岐の条件, 真の場合, 偽の場合)

分岐の条件 ここに書いた式の値で、枝分かれする。0は偽（条件を満たさない）と判定される

真の場合 条件を満たしたら、ここに書いたものを返す。式を書いてもOK

偽の場合 満たさない場合は、ここに書いたものを返す。式を書いてもOK

12に等しい場合は、そのまま。それ以外は1を引く、と考えて作れば

J10

=IF(I10=12, H10, H10-1)

となる。等しいかどうかの判定には「==」を使うプログラム言語が多いが、Excelの式では、等号は1つだけでよい。等しくない、はExcelでは「<>」を使う。上の式を逆の視点で書けば、

J10

=IF(I10<>12, H10-1, H10)

となる。どちらでも、正しく動作するので、考えついた方を入れればよい。

さて、今回は12月から2月まで、を一区切りとしたが、今後もし、11月からとか、あるいは3月までのデータを使うとしたら、このままでは正しく動作しなくなってしまう。また、経済や教育などさまざまな分析シーンで、年度単位でデータをまとめたいときだってあるだろう。

「月」が1~3なら「年」から1をマイナスする。それ以外は「年」そのまま、という枝分かれ処理を式に書けば

J10 =IF(I10>3, H10, H10-1)

となる。

これならば、暦年から年度に換算するように働く、一般性の高い式になる。今回のケースにも対応できる。

②時刻型に変換する

見た目が時刻の文字列を、本当に時刻型にしてやるには、「TIMEVALUE」を使う。日中のデータや夜間のデータに注目したい、というニーズがあるだろうから、ただの文字列ではなく、ちゃんと時刻に直しておきたい。文字列だと、辞書順では1の次が10、次が11……とソートされることになるが、時刻なら1時の次は2時、その次は3時だ。

K10 =TIMEVALUE(D10)

このとき、小数になってしまったら、それは日付や時刻を表す数値（シリアル値）の正体がそのまま見えてしまっている。その場合は、列の表示形式を「標準」や「数値」から、「時刻」に変えてやる。秒まで表示されて煩わしかったら、やはり表示形式のところで「その他の表示形式」>>「時刻」>>「13:30」を選ぶと、分単位の表示になる。

③曜日も知りたい

気象データは、気象庁のサイトからダウンロードする際に指定すれば曜日を含めることもできる。ただし、今回は入っていない。Excelに計算させて、曜日を表示する列を作ろう。いくつか方法はあるが、日付データの見た目だけを曜日表示に変えることにする。「TEXT」関数を使う。

見え方の指定を「"AAA"」にすると漢字一文字で「土」、「"AAAA"」だと「土曜日」のスタイルになる。

L10 =TEXT(G10, "AAAA")

Aを3つにするか、4つにするかは、お好みで。

3.7. 日時のマッチング

①VLOOKUPを使うための準備

これで電力にもキー列ができた。ようやく気象データとマッチングさせられる。

VLOOKUPを使うのだが、検索する先の左端がキー列でないとダメ、という仕様だった。気象データは「JMA」という名前のついたテーブルになっている（「**Ctrl + F3**」で確かめてみよう）が、キーとなる日時は左から2列目で、残念ながら左端ではない。手っ取り早い解決法は、左端のA列を丸ごと削除してしまうことだ。読み込み時にクエリーで削除するやり方もある（更新を繰り返すなら、こちら）。これで、「JMA」という名前がVLOOKUPで使えるようになる。気象データのテーブルの2列目が左端になるように範囲指定して、自分で名前をつけてもよい。

②VLOOKUPの式を入力

M10セルに、仮に次の式を入れてみよう（後で別の式で上書きする）。どんな指示か、意味は分かるだろうか？

M10
=VLOOKUP(F10, JMA, 1, FALSE)

電力データの日時と同じものを、気象データ（または、JMAと名前のついている検索範囲）の左端列から探して、一致したら同じ行のデータを返す、ということだ。行数を1としているので、起算点のままであることに注意。2なら1行右、3なら2行右という指示になる。

③SUBSTITUTE関数による置換

先回りして、マッチング相手の気象データがない時期の電力データは削っておいたはず。なのに、まだエラーが出ている。マッチしない行がちらほらあるようだ。

そう、日付の変わる深夜0時が、よく見ると気象データでは「00:」となっているのに対し、電力では「0:」で、相違があるためだ。そこで、0時を00時に置き換えて、揃えることにする。ただし、「**Ctrl + H**」による置換ではなく、同様の働きをするSUBSTITUTE関数を使おう。でないと、元データが更新になるたびに、手作業しなくてはならないからだ。

▶SUBSTITUTEによる置換

=SUBSTITUTE(操作対象, 検索文字列, 置換文字列)

操作対象 ここに書いたものを対象に、検索と置換をする

検索文字列 この文字列を探す

置換文字列 これに置き換える。空っぽの文字列（""）を指定すれば、削除として働く

という文法。いくつも検索文字列が見つかった場合に、全部を置き換えるか、何番目だけを置き換えるか、というオプションもあるが、省略。

先ほどの式を上書きすることにして、

M10
=SUBSTITUTE(F10, " 0:", " 00:")

と入れる。見にくいかもしれないので、半角スペースの代わりにXを入れると、

=SUBSTITUTE(F10, "X0:", "X00:") ※Xは半角スペースに打ち替えて！

だ。これで、F列にあるデータのうち、半角スペース + 0 + コロン、という並びになっている箇所を見つけたら、0ではなく00にする、という指示になる。半角スペースとコロンで挟むことで、分

や秒や月ではなく、時間が0で始まっていて、かつ00ではないとき、と限定できる。過不足なく、全てを捕まえられる。

なお、コロンはなくてもこの場合は大丈夫なのだが、半角スペースを入れ忘れると、全然違う動作になってしまうことに注意。10時でも、00分でも「0:」には合致してしまうため、「10:00:00」が「100:000:00」に化けてしまう。

もし、列全部が修正されないときには、さっきの式が残ってしまっている。M10セルの右下隅をダブルクリックして、下まで式を書き換えさせよう。

④F列ではなくM列をキーに使う

F列も、いま作ったM列も一意だが、気象のキー列とスタイルが完全に同じなのは、G列のほうだ。こちらを電力のキー列として使う。できれば、タイプして打ち込むようにして、列名（英字のほう）のみ絶対参照にしておく。「F4」キーを1回押すだけだと、行（数字のほう）にも\$マークがついて固定になってしまうので、さらに押して、列だけ固定、行はスライド、にしておく。理由はすぐに分かる。

N10

=VLOOKUP(\$G10, JMA, 1, FALSE)

となる。これで、すべてがマッチするようになったはずだ。ただ、12月1日0時のみは、電力データはあるが気象データはない、ということが分かる（最初にそれぞれのデータを観察したときに先刻承知かもしれないが）。

3.8 気温と降水を結合

ところで、なぜ、上で列数を「1」にして、マッチしたものの自身を取ってきたか（右にずらさなかったか）といえ、意図した通りに捕まえられているかを、確認したかったからだ。電力側のキーをM列ではなくF列の内容で突き進んでいたとしても、つまり00時のケアをしないままでマッチングを試みていたとしても、日時どうしを突き合わせるこのステップを入れておくことで、異変に気がついたはずだ。

意図した通りに動いていることが確かめられたので、いよいよ、気温や降水を取ってこよう。

①気温を取得する

N10のセルに入れた式を、セルのコピーやドラッグで、そのままO10とP10のセルに複写する。少し手直しするだけで使えるからだ。

まず、キー列は日時で、絶対参照になっていればG列のまま変わらない。それで、列方向にスライドしないよう、先ほど絶対参照にしておいた、というわけ。検索範囲も変わらない。違うのは、何列右にずれたところを取ってくるか、だけだ。

気象データでは、マッチした日時の列を起点として2列目、つまりすぐ右に気温があるので、列数を2に打ち替えてやるだけでよい。数式が下まで一瞬で変更されるはずだが、変わらないときには、セルの右下隅をダブルクリックして書き換えてやって。

手でタイプする場合は、以下の式になる。

O10

=VLOOKUP(\$G10, JMA, 2, FALSE)

②「降水なし」を取得する

降水量が0ミリでも、降水なしとは限らない。降水量の列とは別に、降雨・降雪の有無のデータが入っている列が用意されている。降水「なし」が1、「あり」が0になっている。

こちらは、日時を起点に6列目。なので、列数を6に手直ししてやる。

P10

=VLOOKUP(\$G10, JMA, 6, FALSE)

これで、マッチング部分は終わり。

ちゃんと取得できているか、フィルターを利用して見てみよう。空欄やエラーしている行はないだろうか。エラーはないはずだが、2019年の12月3日と9日に、気温が欠測の時間が計3回あるはずだ。不要の場合は、その日時のデータを削ること。

3.9. 見やすくする

でも、降水の有無が0と1で表されるのは、分かりにくいので、「あり」「なし」という表現に変えたい。それから、段階を踏んで処理を繰り返してきたために、表がずいぶん横長になってしまっている。元はといえば、気温と電力の関係を見たくて始めたことなので、この2つは近くに配置したい（温度の体感を左右しそうな、降水の有無も）。

①あり・なし、に置き換える

まさか、「Ctrl + H」で手作業しようとは、思わないだろう。では、SUBSTITUTEか。「1」を「なし」に、「0」を「あり」に、2段階の処理をしないとイケないので、面倒だ。フィルターで見ると、この列には「0」と「1」しかないので、IFで分岐処理してやることにした。

Q10

=IF(P10="0", "あり", "なし")

どうして、ゼロが二重引用符で囲まれて"0"になっているか、分かるだろうか。文字列の0であって、数字の0ではないからだ。数字の0と比較すると、ひとつも一致しないので、常時雨が雪が降っている、というおかしいことになってしまう。逆の発想で

Q10

=IF(P10="1", "なし", "あり")

と書いても、もちろんよい。

②気温・降水・電力を近くにまとめる

これは、難しいことは何もない。すでにある列を、そのまま取ってくるだけだからだ。

R10

=O10

S10

=Q10

T10

=E10

とする。

3.10. CSVにする

ここまで来れば、あと少し。データの受け渡しに便利なCSV形式のファイルを出力しておこう。シートを複製して、値だけにし、最終的には不要な作業列と行を削除する。列の見出しで分かりにくいもの・直したいものがあったら修正して、CSVファイルにすれば完了だ。

①値だけのシートを作る

Ctrlドラッグでシートを複製。シート全体を指定してコピーし、「値だけ」の貼り付けをする。

②不要な列を削除

薄い青の見出しの列全部と、紺色の見出しの列全部を残しておけば、年月や曜日、時間帯など切り口を変えていろいろ探索できる。ほかの作業列は削除する。説明を入れてある部分やタイトルなど、不要な行もすべて削除する。残すのは、表の部分のみにする

薄い青と、紺色の見出しの列だけ残し、あとの列は削除する

③Excelファイルを保存

まず、ここまで作業してきたExcelファイルを保存する。上書き保存よりは、適宜名前をつけるのがよいだろうが、そこは好きに。CSVファイルを書き出すのは、その後にする。でないと、作業過程が全部消えてしまう。

④CSVを書き出す

Excelファイルのうち、完成形のシートを選び、「ファイル」>>「名前を付けて保存」と進む。ファイル名は適宜、タイプは「CSV UTF-8 (コンマ区切り)」を選んで、「保存」を押す。

このとき、今見えていたシート以外は廃棄されてしまうし、このシートの書式や計算式もすべてふっ飛んでしまう。なので、先にExcelファイルをきちんと保存しておくこと、が欠かせない。

3.11 探索してみよう

別々に入手した気象と電力のデータを、日時というキーを使って突き合わせ処理することができた。さあ、この冬の電力消費はどうなっているだろう。お住まいの地点の気象データ・契約なさっている電力会社のデータでも試してみたい。

①Excelでも散布図は描ける

ここで処理した結果をFlourishに渡すのではなく、引き続きExcelで処理して散布図にしたものが「3-3)Excel散布図.xlsx」。テーブルがあるシートの「スライサー」のボタンをオン/オフすることで、描画するデータを選べる（テーブルのフィルターと同じ働き）。グラフはその都度、自動的に描き直される。グラフの中の線が、各シーズンの回帰直線だ。

シーズン別に色分けするために、少し手を加えてある。しかし、それはExcelの仕様のせいで、こだわらなければ、もっと楽に作れる。Flourishで散布図の雛形を選び、いろいろ試してみると、おもしろいと思う。

4. 文字情報がなく画像だったら

きょうここまでは、取り込みたいデータが文字情報になっていること、が前提だった。

文字情報を持っていないPDFや画像形式のファイルから、文字を抜き出すにはどうしたらよいか。でも、間違いなく時間が足りないので、各自で試してみて。

Google Driveを使う

Google Drive (<https://drive.google.com/>) にファイルを保存。右クリックして「Google ドキュメントで開く」を選ぶ。画像部分だけにしたPNGファイルをアップロードしたら、あっという間にテキストに変換された。ただ、私が試したケースでは、ひらがなの「へ」とカタカナの「へ」はどちらもひらがな、カタカナの「タ」と漢字の「夕」はどちらもカタカナと認識された。

なお、文字情報と、一見すると文字に見える画像部分の両方を含むPDFファイルでも試してみるところ、画像部分はテキストに変換されないままだった。こういう場合は、画像だけを保存してやらないと、だめらしい。

参考：Tabulaを使う

こちらは、文字情報を持っているPDF限定。表形式のデータの取り込みに威力を発揮する。

<https://tabula.technology/>

参考：Wordを使う

Excelから直接、PDFファイルを読み込めないときに役立つ。ファイルを右クリックし、「プログラムから開く」でWordを選ぶ（Wordのバージョンによっては非対応。Office365の場合は大丈夫）。表形式を保ったままExcelに取り込みたいときに、Wordで仲介するとうまくいくことがある。

5. 付録

能率が上がるExcelショートカット

▶Ctrl + ドラッグ

シート名のタブをつかみながらドラッグすると、そのシートが複製される

▶Shift + ドラッグ

行や列を選択し、その境目をつかみながらだと、並び替え

▶Ctrl + 1

セルの書式設定。エルではなくて数字の一（テンキーの1はダメ）

▶ **Ctrl + Z**

直前の変更を元に戻す

▶ **Ctrl + A**

まとまった範囲を選択。もう一度押すか、何もない部分で押すと、シート全体を選択

▶ **Ctrl + ***

データが入っている範囲を選択。離れ小島は選択されない

▶ **Shift + カーソルキー**

選択範囲の拡大・縮小

▶ **Ctrl + カーソルキー**

空白セルは飛ばし、その次にデータの入っているセルにジャンプ

▶ **Ctrl + C**

コピー

▶ **Ctrl + V**

通常の貼り付け。セル幅以外すべて引き継ぐ。Ctrlだけをもう一度押すと、貼り付けの形式を選ぶ

▶ **Alt + Ctrl + V**

形式を選択して貼り付け。Ctrl + Vでは困るときに使う。関数を使って整形をした後、貼り直して「値だけ」にするのに便利（単独でVを押し、さらにEnter）

▶ **Alt + ;**

絞り込み時に表示されているセルだけをコピー元にする

▶ **Ctrl + S**

ファイルを上書き保存

▶ **「F12」**

ファイルを別名で保存

▶ **「F2」**

セルの編集。数式を入れるときに便利。

▶ **Ctrl + F**

検索

▶ **Ctrl + H**

置換

▶ **Ctrl + Home**

A1セル（左上）にジャンプ

▶ **Ctrl + End**

データの入っている最終セルにジャンプ

▶ **Ctrl + PgDn または PgUp**

次のシートや前のシートに切り替える

▶Ctrl + ;

きょうの日付を入力。便利

▶Ctrl + :

現在の時刻を入力。便利

▶Ctrl + T

テーブルにする

▶Ctrl + Enter

複数のセルに同じデータを入れる。一括して修正するときに便利

▶Ctrl + F3

名前の管理

▶Alt + 下矢印

そのカラムに入力済みのデータのリストから選ぶ

▶範囲選択した後、Alt + H, O, I

列の幅の自動調整。あらかじめCtrl + Aなどで範囲を選んでおく

回帰直線とは

- Y列の値をX列の1次式で説明・予測しようとするもの。直線で近似するのが妥当なときしか、やる意味がない。名前はいかめしいし、回帰ってどういうことよ、と思うが、そういうこと
- なので、直線的な関係がありそうか、先に散布図を描いて見当をつけるべし
- 直線的な関係があるとき、という以外にも制約はあるが、省略
- X列とY列の平均値を通り、かつ、実際の値とのY軸方向のズレを最小する直線を引く。あくまで、X列でY列をどのくらい説明できるか、に使う（今回であれば、気温が電力消費を左右する、という方向。電力が気温を決める、は変でしょ）

6. 終わりに

本日は、長時間お付き合いありがとうございました。

終了後も、質問はご遠慮なくお願いします。ひとりが分からなかった部分は、多くの方が困った難所のはずです。復習してみて詰まったときも、Slackでお尋ね下さい。私が分からないところもたくさんありますが、きっと仲間や先達が助けてくれます。

7. 付記

ファイル読み込みのクエリーで何ができるか試してみたら、かなり自動化できて、読み込み・貼り付け・JMAデータの1列目の削除といった手作業を完全に省略できることが分かった。ただ、クエリーの文法がまだ飲み込めない上、接続先を相対パス指定ができるのかも、私には分からない（ほかの方のPCでは動かなくなるかもしれない）。いろいろ試して、いつかご報告できればと思う。お詳しい方は教えて下さい。

Excelファイル本体を軽量化できる上、不要な行の削除を気温の列のフィルターでやる（#N/A行と空白行）だけで、ごく短時間で処理が終わるのは、データの更新を繰り返す場合に助かる。行の削除も、やり方次第で自動化できるのではないか、という気がする。

データの出所

使わせていただいたデータの出所は以下の通り。

国連決議

- ・2014年3月27日分

<https://digitallibrary.un.org/record/767565>

- ・2022年3月2日分

<https://digitallibrary.un.org/record/3959039>

- ・2022年10月12日分＝PDFファイルでこのレポにあり

気温

- ・気象庁

<https://www.data.jma.go.jp/risk/obsdl/index.php>

上記から、地点と項目を選んで、3カ月分程度ダウンロード。それを繰り返す。

電力

- ・東京電力

<https://www.tepco.co.jp/forecast/>

「データのダウンロード」から、2018年以降の分のCSVファイルをダウンロード。

直近分は詳細なデータがある半面、まとめて1年分をダウンロードすることができない。でも、古いデータをダウンロードしたときのURLを見てみると、4桁の数字が西暦年を表しているようだ。その部分（下記のXXXXのところ）を2022や2023に置き換えて、やってみた。

<https://www.tepco.co.jp/forecast/html/images/juyo-XXXX.csv>

とはいえ、ハンズオンではなく報道目的ならば、まずは取材して正面からお願いするのが筋。

そのほか

テーブル内のデータを行単位で削除するために、右クリックから「削除」を選んだときの、次の選択肢と挙動が謎。フィルターで絞り込みしていないときは「テーブルの列」と「テーブルの行」なのに対し、絞り込み時は「テーブルの列」「シート行全体」になる。「シート行全体」のときは、確かにテーブルの範囲外も丸ごと削除されている。

フィルタリングしているか否かで挙動が変わる必然性はないように思うが、どんな理由があるのだろうか。テーブルの中だけで完結する削除方法はないのだろうか。

また、絞り込み表示のときに、表の中ではなく、画面左端の行番号のところで範囲指定して複数行を削ると、フィルタリングで非表示になっている行まで、ごっそり削られてしまった。Alt + ;で、見えている部分だけにできないのか試してみたが、うまくいかない。

どなたかご教示いただけましたら幸いです。