\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Migration/Refugees -- Data Module  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

***1.1: Learning Objectives***

Good visualizations are manipulative. And being *manipulative* isn’t always bad. What if it means giving voice to the voiceless? Missing Migrants may be missing data, but does that mean they should be missing from visualizations, too?

Migration data is vulnerable to manipulative visualizations. But how we choose to deal with that vulnerability is a separate issue. We anticipate that students following our process for migration data visualizations will have a common issue: many migration data sets are incomplete--as such, part of our data module’s goal is to show how incomplete data sets can still be used for data visualizations, but also show common areas that the incompleteness can lead to misleading outcomes. In that case, we hope to offer some workaround solutions (some of them more elegant than others) to help students in their own future visualizations.

Our data management practices (and those modeled in our module) will focus on the selection of data for visualization. The goal of our data management section is to simulate how different data visualizations can look in their outcomes--whether they are intentionally manipulated, or mistakenly misleading. The goal here is not to show step-for-step how to perform data management--we anticipate that will be covered throughout the various modules in a variety of ways. Instead, we will focus more on the research design elements and analytical elements--primarily, how we set up our narrative workflow (we call it the “Preparation & Manipulation” workflow), and the primary areas where our biases creep in (and to determine whether we want them to or not).

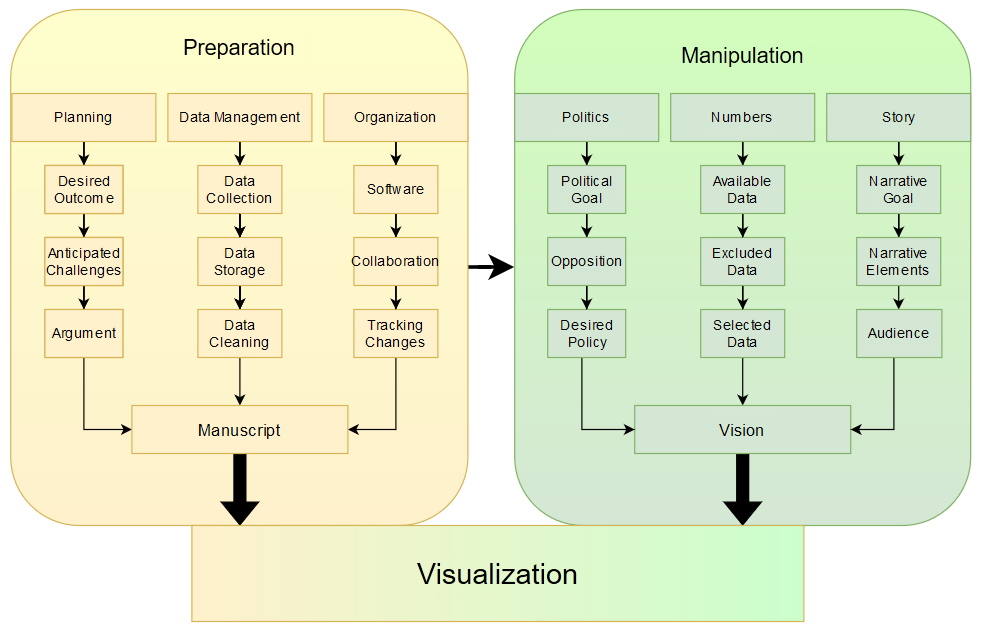
Data management fits into our overall story as the staging ground -- what preparations are necessary, even at the most basic level, to be able to get data (and maintain or clean it) in a way that better allows you to manipulate it in the future. This is a critical step in the manipulation process. Yes, data sets can be manipulated through visualization alone, but the concept of “look at what you can do when you have the power to manipulate the data yourself!” could be an eye-opener for students--especially when they think about the “narrative” behind visualizations.

For this module, we will be using Tableau Public, a data visualization software. Many tools exist for data visualization, and they each have their own strengths and weaknesses. Data viz software like Tableau have some advantages in usability and accessibility--the “skill floor” is lower. Programming languages give users free rein to create within the limitations of their language, but they also require a deeper familiarity with the software. In other words, they aren’t “plug and play.”

Visualization software programs like Tableau allow users to create complex data visualizations with just a few clicks--but they have their cons, too. The more complex the visualization (and the less you understand about the software), the more likely you are to make a mistake without realizing (or understanding) it. You might select the wrong variable, observation, or even the wrong data. The less familiar you are with the software, the less familiar you will be with its presentation. To avoid that, we’ve created a workflow that starts by having you familiarize yourself with your *data.* If you’re familiar with the data, you’ll be more comfortable noticing errors in your visualization. Later on, when we get to the visualization stage of the exercises, we’ll talk about some of the tools Tableau has to help inexperienced users with the data cleaning process.

Before we get into the exercises, let’s look at the process for familiarizing ourselves with the data. And remember: these are guidelines, not hard and fast rules--your exact steps may deviate, but your *process* should look similar.

***1.2: Data Workflow***



***1.3: Detailed explanations of all the steps in the workflow.***

The workflow begins with a two-stage framework: Preparation and Manipulation. The framework has a singular goal: creating a *manipulated outcome*. The *manipulated outcome* is your story--it’s the result of both *what* you want to tell and *how* you tell it. In this module, that outcome will ultimately be achieved through a visualization, but the preparatory steps that guide the visualization process are equally as important. Preparation is the foundation of any good story, and the same can be said for good visualizations. These steps are not written in stone--they are simply a guiding framework to help you think about the story behind the visualization process.

*Preparation*

ThePreparation stage is just that--the foundation. It’s the step where you think about the data and logistical elements of your visualization. The Preparation stage is broken into three foundational steps: planning, management, and organization. If the final *manipulated outcome* is what the reader takes away from your story, think of Preparation as your brainstorming session. You want to break up the planning process into “manageable chunks[[1]](#footnote-0).” In other words, organize your decision making.[[2]](#footnote-1) You can’t expect to create a successful visualization by luck alone. Instead, first plan out your objectives, anticipate any challenges, and decide what your ideal outcome would be--what do you want someone to take away from your story?

The Planning stage is the first (and most critical) step when building your outcome--it’s the foundation of your foundation. Here, you’ll first think about your desired outcome--what is your goal? A compelling visual for a handout? An evocative image for an otherwise boring slideshow? A beautifully clean visualization of complex data?

Unfortunately, there will always be a problem--perhaps the data that you have feels old, poorly organized, or incomplete. Perhaps you do. Either way, Anticipating Challenges before you begin your work will help you craft a more compelling story.

Another important consideration for telling a compelling story is your Argument. When you think about your argument, consider the *manipulated outcome* as a whole--when someone sees your visualization, what will be their response? Think about the different kinds of responses you can create--will you appeal to the viewers needs, emotion, or logic? For narratives built around human data--for example, data on missing migrants, creating a visualization that appeals to the emotions of viewers, rather than their logic, may have a stronger effect on the audience.

The Data Management stage is for collecting, storing, and cleaning your data. Data management fits into your overall story as the staging ground. Properly managed data, even at the most basic level, allows you to better manipulate it in the future. Data can be manipulated through the visualization process alone, but when you collect, store, and clean the data yourself, your options for successful manipulation increase. For our first exercise, we’ll be using data sets from KNOMAD on bilateral migration and remittances--some data sets will require light cleaning, and we’ll also look at what happens to your visualization if you don’t clean your data set properly. Spoiler alert: it might be completely wrong.

The data management process may also involve making your data *tidy.* Hadley Wickham, a statistician from New Zealand, describes tidy data sets like families: they’re all alike, while every messy data set is messy in its own way.[[3]](#footnote-2) The fundamental reason for making data sets tidy, or standardized, is to make them easier to work with, visualize, manipulate, and model.[[4]](#footnote-3) To make data “tidy,” data should be arranged so that each variable is a column and each observation a row.[[5]](#footnote-4) Tidying data is one way to organize data, but it’s not the *only* way to organize data. For more information on the tidying process, refer to Chapters X and Y (or further readings, tbd.)

After we’ve organized our data, we need to organize our *project.* The Organization stage is about deciding what software, storage, and collaborative tools to use. When deciding what visualization software to use, consider not only what software is available to you in terms of access, but also understanding--some students may be very comfortable with online visualization software or Excel, while others may prefer to write their own code. In this module, we’ll be using Tableau. Our goal is to provide students of all levels of access and expertise with some basic visualization skills.

For collaboration and storage, Google Drive, Google Docs, or another form of cloud storage may be sufficient for your project. In other cases, you may want to upload your data to a repository like GitHub or OSF (Open Science Framework). For the sake of this exercise, you’ll be working with data previously uploaded to an OSF repository. For more information on uploading and storing data on GitHub and OSF, refer to our “Additional Reading” section at the end of the chapter.

The combined efforts of your Preparation stage will create your Manuscript. This is the backbone of your process. It doesn’t have to be a sprawling document of exhaustive notes and planning--it can be as simple or as complex as you need it to be. Good documentation during the preparation stage will, however, make it easier to catch problems *before* they become problems. If you throw some data into a visualization without properly cleaning or storing it, you may run into problems late into your efforts--problems that may require you to start the process over.

*Manipulation*

The next stage in the framework is Manipulation. The word “manipulation” can carry a negative connotation--one of psychological subversion and secrecy. Unfortunately, that isn’t what this stage is about. We want to distinguish “*manipulation*” from “*manipulative.*” The outcome, your *manipulated visualization*, doesn’t require you to undermine your audience with a cleverly misleading visualization. That being said, it certainly can*.* A *manipulative* visualization may very well be effective, and at times, even appropriate.

In a textbook about the politics of numbers, it should be no surprise to see their inclusion in our framework. The Politics stage is where you will analyze the political (or policy) goals of your manipulation. Do you have a primary political goal--perhaps related to immigration, withdrawing a union, or both? Consider the major oppositions to your goal politically. If your agenda is for immigration reform, you may consider a visualization or set of visualizations that preempts some of the major opposition arguments. Lastly, think about your policy goals. What policy or policies are you hoping to influence (or outright change) with your visualization? This is the stage where you focus on your concrete goals, rather than the abstract.

Numbers and concrete goals--name a more iconic duo. In this stage, you’ll examine the available data and data sets, determine what data you should exclude (if any), and select the right data for your visualization. A successful Preparation stage should make this part easier, but there are still some final decisions to be made. Those decisions, however, may change as you create your visualization, and that’s okay--because you’ve already prepared and managed your data, you can manipulate it as you see fit.

Story time. The last step before working on your visualization focuses on finding your Story. In this stage, you’ll think about your narrative goal, visual elements, and audience. Your audience is critically important when creating your visualization, but for now, focus on the big picture: what story am I telling? What is the tone?

The final (and most fun) stage is all about the Visualization. This section is where you may spend the majority of your time and effort--creating a compelling visualization might involve some very minor manipulations (but lots of them). It might also involve some major exclusions. Some of these changes might be simply visual changes--making the font larger here, a bubble larger there--while others may focus on the final (and very critical) step of this stage: the audience.

In your academic and professional life, your audiences will vary. Some people appreciate a nice story. Some people want the facts. Sometimes they’re the same people--usually they aren’t. Those different audiences exist in policy and politics, too. It’s critically important to consider your audience, and later questions in this module will focus purely on that aspect. Who are you trying to influence (or manipulate), and how can you most effectively do it? Oftentimes, your audience will be unfamiliar--that’s fine. Your goal is to make key assumptions, and use those assumptions to make a compelling case (or at least a compelling visualization).

After a successful Manipulation stage, you will have created your Visualization. This should be representative of your overarching goals: an amalgamation of your policy aims, your selected data, and your vision. Pat yourself on the back, and try not to think too hard about how much work that was for some numbers in a picture.

*Manipulated Outcome*

The final product. The combined efforts of your preparation, manipulation, manuscript, and visualization. The outcome may or may feel like a success, but “success” was never the ultimate goal--the goal was to create the *most compelling* story, told through both the preparation and visualization, and hope that it has the *most effective* outcome*.*

***2.1: Learning Objectives***

In this data module, our goal is to guide students through the storytelling side of data visualizations. As you’ve learned from the chapters, no data is without its story. The same can be said about visualizations. This module will help students think about how to use data, even basic Excel data, to tell a story through visualization--*any* story. Data manipulation doesn’t always occur in overt, or even intentional, ways. A missing (or excluded) value may tell a misleading story--but it might also tell a more accurate one. Sometimes, manipulation may occur to intentionally mislead, or create opportunities for corruption.[[6]](#footnote-5) As data scientists, our aim (hopefully) is to use, rather than misuse, the story elements of data--especially the narrative power of visualizations. They’re analytical shortcuts--whether we want them to be or not--and we like them.

Actually, we love them. We cut desire paths through fields to save ourselves seconds. Someone reading this textbook wrote an entire paper on a book they never read--and it was *good*. To some people, “conclusion” is synonymous with “summary.” Humans love shortcuts. And our governments, our politicians, our data scientists, and our toothpaste companies know it. We encounter manipulated data every day. Whether it’s a commercial, a press release, or a migration report, you’ve probably asked yourself some variation of the following question: “Why did they include *that?”* We hope to help you better answer that question--or, more accurately, *understand* the answer.

As data consumers, collectors, and creators, our best defense against manipulation is awareness. And although the sentiment may sound like pre-apocalyptic government propaganda, it’s true. In a perfect world, data would always be used to tell its “most accurate" story. But we don’t live in that world (yet)--there is always a gray area. Even “most accurate” itself is a gray area--most accurate to whom? The data, the producers, the donors, or the surveyed populations? And where do data deprivations and missing data fit in? In this module, we want to look behind the curtain of our imperfect world and understand how “most accurate” may not always mean “least manipulated.”

Other modules in the textbook may focus on the different stages of data management: data cleaning or tidying, organizational methods, or even writing your own custom code for visualization. But we recognize that not all students are experts in these areas--neither are we. But that’s the beauty of manipulation--it’s primarily concerned with being *effective*. And effective visualizations (whether intentionally manipulated or not) are not always written in custom code, well organized, or properly tidied--sometimes, that may even be their strength.

Within this module, we’ll examine how those steps can be used (or misused) to manipulate data. We’ll examine sample visualizations, examine their strengths and flaws, and figure out how to “course-correct” towards *less* manipulative (or misleading) data viz along the way.

To better understand how these types of data viz manipulations can occur, we break down the narrative process into two steps: Preparation and Manipulation. Follow along with us as we select, prepare, and create basic visualizations--this time using the Tableau software, but our core concepts and objectives should align with any software or methods.

In conclusion (or if you just want the shortcut), internalize this: data can be manipulated without being manipulative, and data can be manipulative without having been manipulated.

***2.2 The Problem Sets***

Our first exercises will be examining the KNOMAD’s (the Global Knowledge Partnership on Migration and Development) compiled remittance data sets. These will be primarily introductory. For the first exercise, we’ll go over some of the answers together. In the next exercise, you’ll be asked to provide your own.

Later, we’ll work with KNOMAD’s bilateral migration data set to create visualizations to match real life political scenarios (and under political constraints)--what if your data and your policy goals want to tell very different stories?

We’ll start with some basic questions with guided answers to get you thinking critically about data and visualizations--even very simple ones. After a brief introduction, we’ll jump right into creating some of the visualizations used for the problem sets, and finally, be given a political scenario (where you can work alone or with a group) to produce your own *intentionally manipulative* visualizations.

**2.2 Exercise #1.1**

Before we download the data sets ourselves, let’s look at some snapshots of the data sets. Our goal is to get into the mindset of someone in the “Data Collection” stage--will this data be suitable for our visualization? Are there any glaring problems or major weaknesses to consider (or exploit)? What else should we consider before we use this data? We’ll start with a snapshot of *Remittance outflow data from 1980-2020,* compiled by KNOMAD, an open global migration partnership.

**Question 1: What are some of the first questions you might have when looking at the data set in Figure 1.1?**

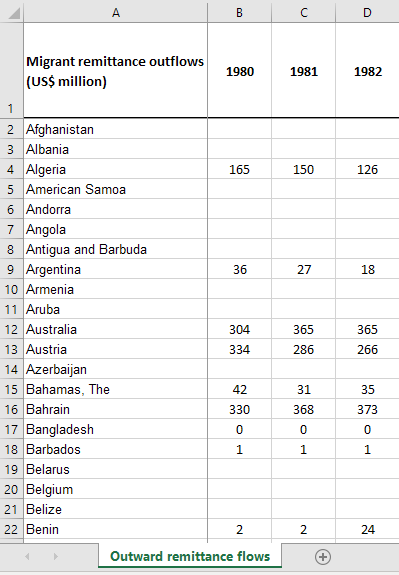


Fig1.1

Our Questions:

1. There are large data gaps or missing data--why?
2. Who are the producers of this data?
3. Is there a difference between no data (null) and the data for Bangladesh (0)
4. Are “remittances” defined anywhere?
5. Do we know *where* the outflows are going?
6. There is a huge jump in Benin’s data from 1981-1982--will jumps like this affect the data set or our visualization?
7. Can we make this data “tidy”?

Our Answers:

1. Remittances were not tracked prior to the early 2000s--*many* countries are missing remittance data.
2. In this case, this info is listed on the spreadsheet: IMF data, central bank data, statistical agencies, country records, and World Bank desks.
3. Yes. Null data is unreported data.
4. Yes. This is defined at the bottom of the spreadsheet.
5. In this data set, no.
6. It could--in some cases, that might be a bad thing. In other cases, it might be useful.
7. It’s fairly “tidy” as it is, but it could be made tidier.

**Question 2: You’ve decided to use the data set in Figure 1.1 for your visualization, what are some of the next steps you should take?**

Potential Answers:

1. Store the data locally on your computer.
2. Check if “remittance” is defined.
3. Determine the producers of the data.
4. Determine if “0” really means 0--is there rounded data in those cells?

Now that we’ve asked (and answered) some of our questions about the data, let’s look at how we might visualize it. For the next exercise, we’ll present you with some “packaged” visualizations. They won’t be perfect--but that’s the point. Let’s see how we could improve upon these visualizations, examine their strengths and weaknesses, and think about what might be a more *effective* visualization given the data.

**2.2 Exercise #1.2**

This exercise is about examining pre-made visualizations, critiquing them, and then creating your own (better) version. We’re going to use the data from Exercise 1.1 to create these visualizations. After we examine the visualizations, we’ll guide you through our process and help you create your own. These visualizations will have their pros and cons--your job is to figure out what is working and what isn’t.

The visualization below (Figure 1.2.1) comes from KNOMAD’s data on remittance inflows. The goal of this visualization was to determine what countries were receiving the most remittances in 2020. The image further below (Figure 1.2.2) is a similar visualization with data from 2020.

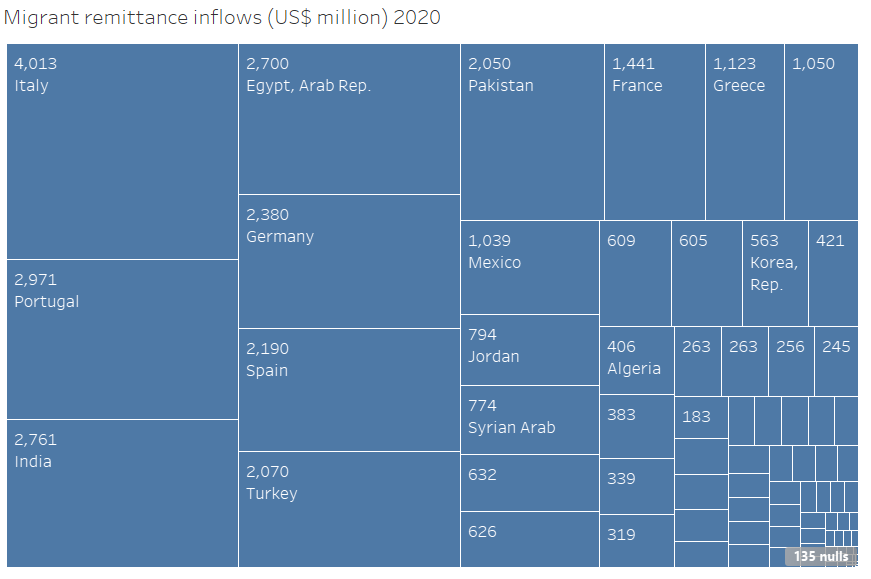
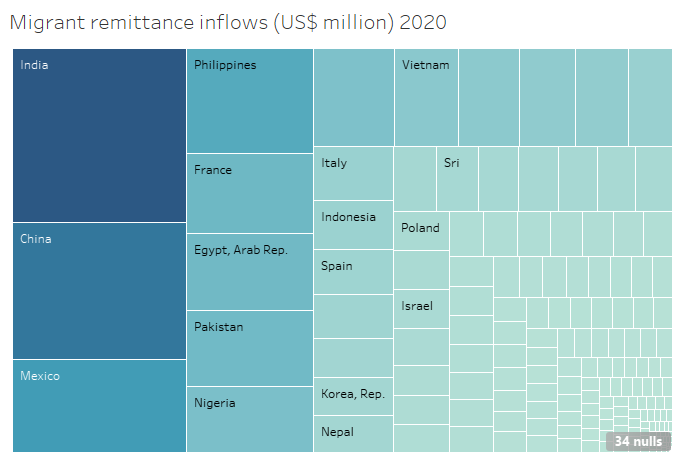


Fig. 1.2.1

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

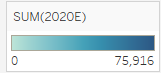


Figure 1.2.2

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Question 1: Compare Figures 1.2.1 and 1.2.2--what differences do you notice immediately? What are your initial thoughts and questions? Remember: the goal of this visualization was to show which countries were receiving the most remittances in 2020.**

Potential Answers:

1. Figure 1.2.1 is hard to read, but the detailed values are useful.
2. Some of the countries are unlabeled--some only appear as cell boxes.
3. Both have significant null values--should we do something about that?
4. With the extra data, tons of countries appear too small to even label--it hurts the aesthetic of the visualization.
5. The gradient seems “top heavy”-- more than half of the viz is the same color.
6. Are the colors used appropriate for accessibility?
7. Is the ‘gradient scale’ clear enough?
8. Should we include values within each cell for readability?

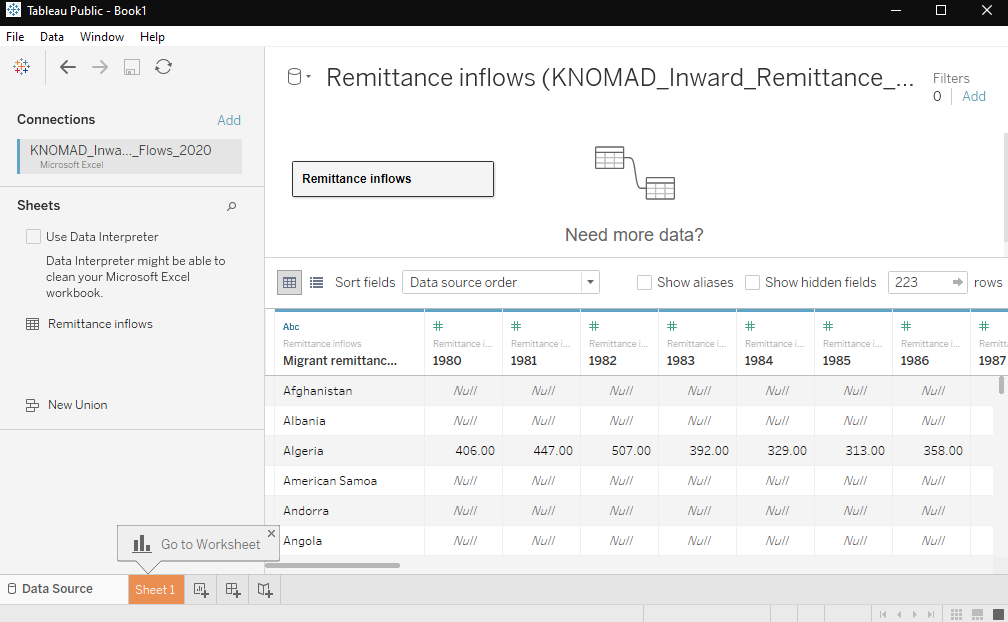
**2.2 Exercise #1.3**

First, let’s reproduce one of our earlier visualizations from KNOMAD’s remittance inflow data set. We’ll go over the process of saving our data, importing it into our visualization software, cleaning it (if necessary), and learn some basic ways to manipulate the data within the program. At the end of the exercise, you’ll use the data from the *remittance outflow* data set to create your own (similar) visualization.

Steps:

1. Download Tableau Public at [https://public.tableau.com/](https://public.tableau.com/en-us/s/), or type “Tableau Public” into your preferred search engine.
2. From the OSF repository (<https://osf.io/ekdr3/>) download: “*KNOMAD\_Inward\_Remittance\_Flows\_2020.xlsx”*
3. Save this file to a location you will remember--or create a new folder for it.
4. Open Tableau Public.

After you’ve opened Tableau Public, navigate to the top left “File” and select “Open.” Navigate to your “*KNOMAD\_Inward\_Remittance\_Flows\_2020.xlsx*” file, select it, and press Open. Your Tableau Public should now look something like this: (placeholder screenshots)



Before we start to create a visualization, let’s look at the data. Navigate to your *KNOMAD\_Inward\_Remittance\_Flows\_2020.xlsx* file and open it in your preferred spreadsheet software (Excel, Google Sheets, etc.). Remember: we want to familiarize ourselves with the data so that we can look for any glaring errors in the “opening” process.

Our data within Tableau and the spreadsheet seem to match up: in both cases, we have ***Migrant remittance inflows (US$ million),*** listed by ***Year*** (across) and ***Country*** (down). Tableau has automatically labeled our empty cells as *Null,* so it’s done a bit of the cleaning process for us--nice. But this won’t always be the case, as we’ll see when we work on Exercise 1.5.

Now that we have our data imported and checked, we can start creating our visualization. In the bottom left next to “Data Source”, select “Sheet 1”, shown below.

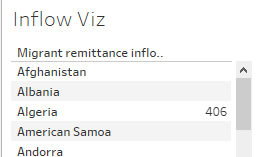


This will be our worksheet for our visualization. Double click Sheet 1 to rename it. For now, let’s call it “Inflow\_Viz”.

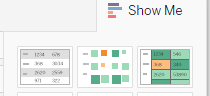


To create the visualization used in Figures 1.2.1 and 1.2.2, we need to select our data. Right now, we have nothing--a blank “worksheet.” Because we’re working with relatively clean data, we can simply double click the data we wish to visualize. In this case, ***Migrant remittance inflows (US$ million)*** and a date, like ***(1980)***. They will then be imported automatically into our sheet.

For now, these figures will be visualized as a text table, like so:



This text table isn’t particularly useful for our large data set. Let’s try something else simple, like a horizontal bar. In the top right, you should have a box called “Show Me”, with various graphs and figures on the right side of your sheet, like so:

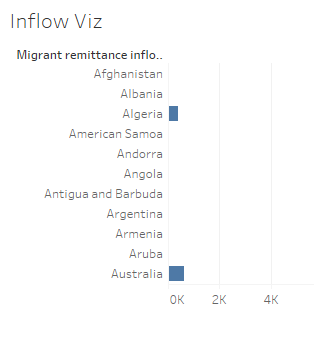


If you don’t see the different visualization options, double click on “Show Me”, and a dropdown menu with the graphs should appear.

Select the “horizontal bar” icon, shown here:



Now, your visualization should look something like this:

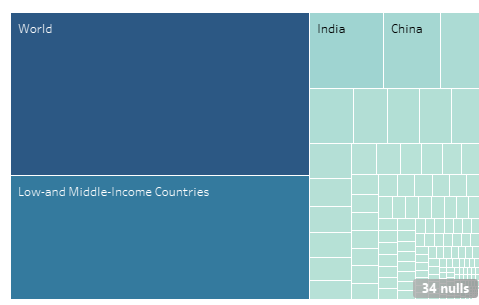


With some cleaning and relabeling of the axes, these horizontal bars might be an acceptable visualization. But will they be *effective* for this large data set? Probably not. Let’s keep looking at some other available visualization options.

For Figures 1.2.1 and 1.2.2, we used a “treemap,” displayed here:

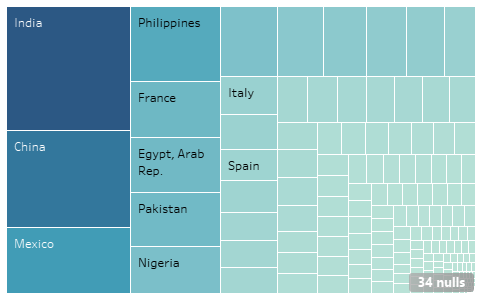
 (Treemap icon)

Now, we should have a visualization that looks something like this:



Right away, you will notice some major flaws. In our data set, we missed something in the cleaning process--two of the rows contained aggregate data. But that aggregate data may be useful later--instead of going back and removing it from the spreadsheet, we can simply remove it from our visualization. Right click the “*World”* portion of the visualization, and select “Exclude.” Now, do the same with the *“Low-and Middle-Income Countries”* section.

Now, your visualization should look something like this:

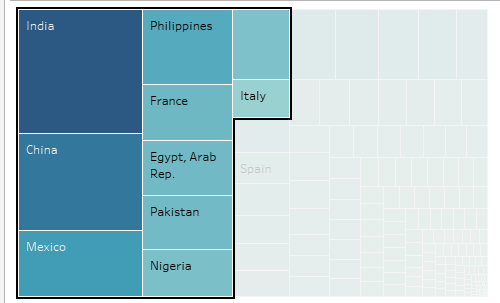
**

It’s starting to look a lot like the visualization from our previous exercise, but it’s not quite there yet. Our title was automatically created from our Sheet name, and it doesn’t really work here. Let’s title it after our variables and observations, maybe something like: *Migrant remittance inflows (US$ million) in 2020*.

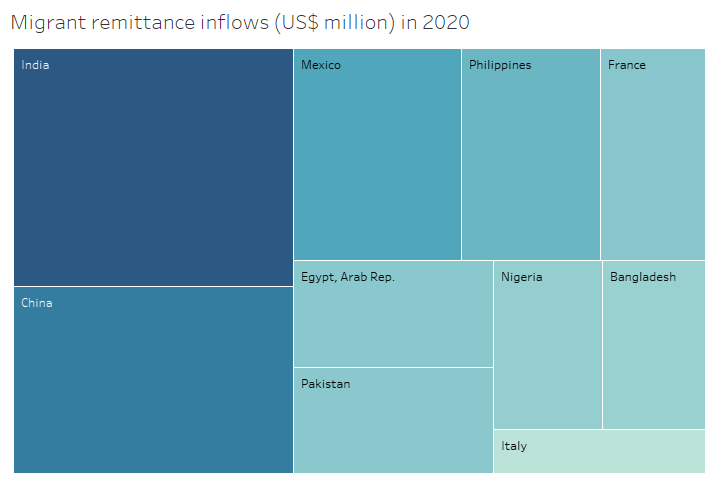
If you remember from earlier, we noted that our data set had a large number of missing or null values--here, we can see them in the bottom right. Even in 2020, this data set still has 34 null values. If we click on that section “34 nulls” we are presented with two options: we can assume all null values are zero, or we could exclude them from the visualization. Depending on your goals, either decision may be effective, but for the sake of being *more accurate*, let’s filter (or remove them) rather than set them at 0. In this case, “no data” doesn’t mean “no remittances.”

Still, looking at our visualization, something isn’t working. There are hundreds of tiny, empty cells with no data and no label--why? Here, it’s mostly because of *scale*. India is receiving an estimated $76 *billion* (75,916,000,000)--compared to the $3 million (3,000,000) Angola is estimated to receive. This is when we need to examine our visualization *goal*--what are we trying to say, or trying to show?

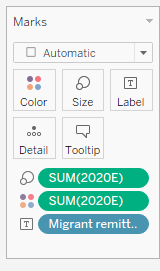
For this example, we wanted to examine the countries receiving the most remittances--let’s say, the “Top 10.” In Tableau, how can we pare down our visualization to show that? We *could* go through and exclude the other 200+ countries remaining from our data set like we did with the “*World”* data, but we value our time and sanity. We could also select by holding Ctrl + left clicking the top ten countries, leaving us with something like this:



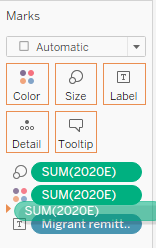
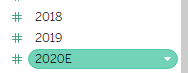
Now, we can right click our selection and say “Keep Only.” This will exclude the remaining countries, leaving us with our top ten. It may take some resizing to get the appropriate labeling, but we can fix that later, if necessary. Our current result should look something like this:



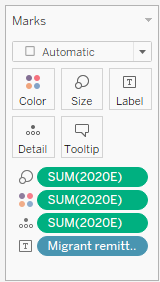
Here, we’re reliant partially on the gradient color scale to determine values (for now, this scale may be hidden under your “Show Me” section, but it can be moved by selecting it and dragging it--you may wish to move it to the left of your visualization for better visibility). In the first visualization--the one without a gradient--you may have liked having the value below the country. To add that to our visualization, we can “drag” the data from 2020 into our visualization again, but this time, we’ll add it as a label.

On the left of your visualization, you should have a box called “Marks,” pictured below:  


This box represents what data we currently have in our visualization. We have the 2020 data categorized by size (size of the boxes), color (the color gradient), and our box titles (country data). But we want to add more information into our visualization--we want our remittance *values* to appear under each country. First, we drag the data from 2020 on the left into our “Marks” box, like so:



Now, our “Marks” box should look like this:

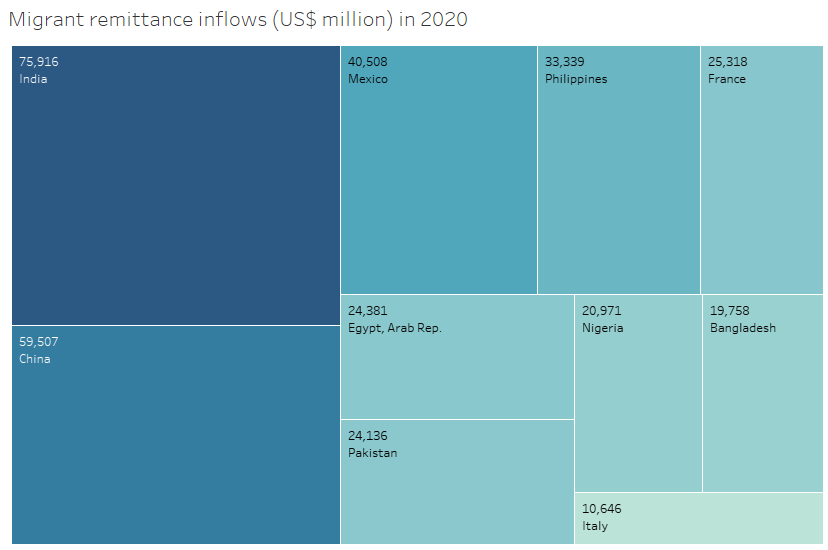


We already have the Size and Color in our visualization, so we’re looking to use this additional data to show a Label. In the image below, we see that it’s currently being used as a “Detail” from the icon on its left:

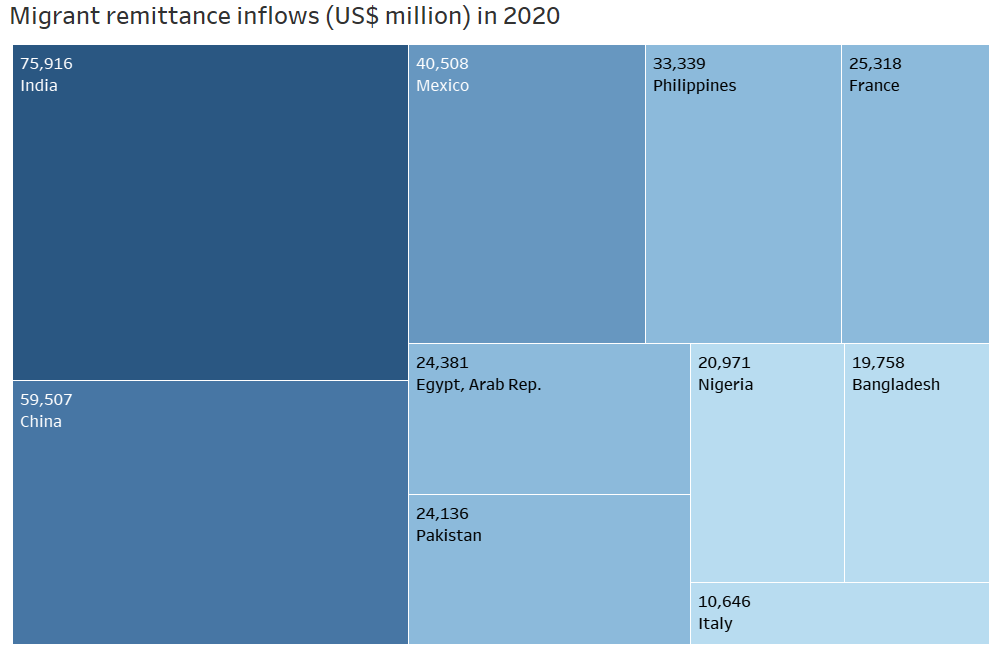




Let’s try turning it into a “Label” by clicking on the icon to its left () and see what happens. The numerical data from 2020 should now be shown on your visualization matching its country, like below:



You can experiment in the Marks box with different colors, font colors and sizes until you have a visualization that you like. You can also experiment with the color gradient scale to make color differences more (or less) noticeable. Before you experiment, you may want to “Duplicate” your worksheet in the bottom left--if something goes wrong, you always have your original to fall back to.



After some fiddling with fonts, colors, and the gradients, you may be left with something you’re happy with. Or maybe not. If not, there’s always coding.

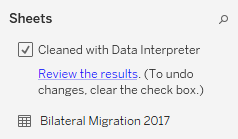
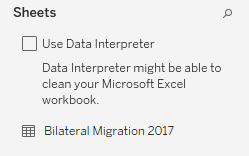
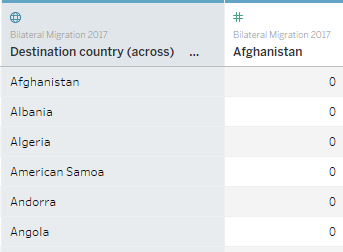
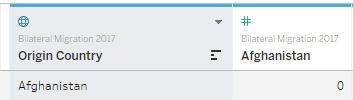
**2.2 Exercise #2.1**

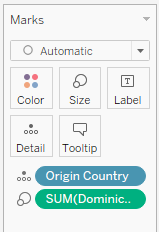
Exercise 2.1 will look at some of the data that matches one of the examples in our chapter--specifically bilateral migration data from Haiti and the Dominican Republic. In this exercise, try and experiment with Tableau to create compelling visualizations to match the following scenarios:  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

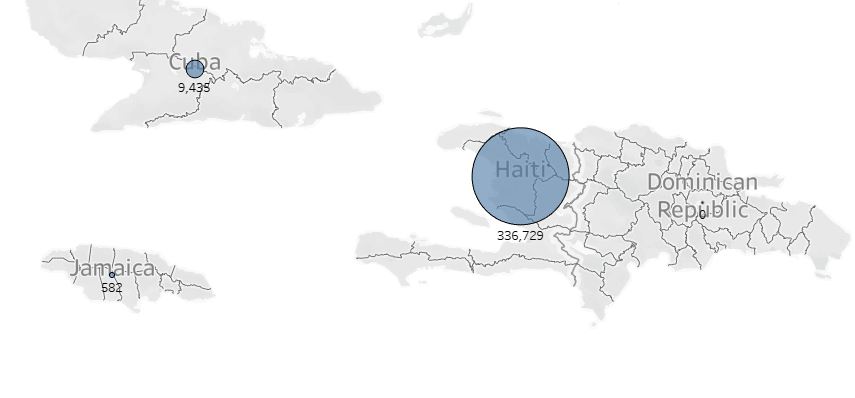
You work for the ruling government of Haiti, and recent disputes over migration at the Dominican Republic border have led to a complete collapse in relations. The Dominican Republic’s government thinks that too many Haitians are migrating to their country--but your government doesn’t agree. Using the data set provided, use Tableau to create a visualization that helps push the narrative that the D.R. isn’t receiving “too many” migrants relative to its neighbors.

Now, imagine the scenario was flipped, and that you work for the ruling party in the Dominican Republic. You want a visualization that shows that *too many migrants* are coming from Haiti--how would you manipulate the *exact same data* to produce a visually different result?

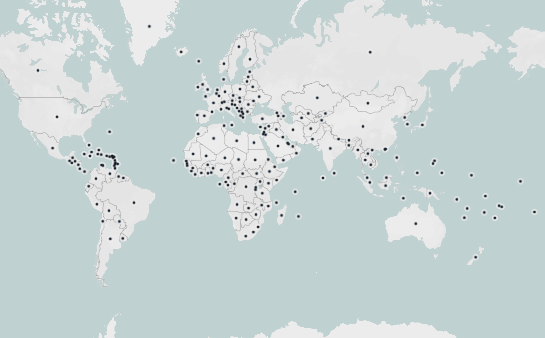
Steps:

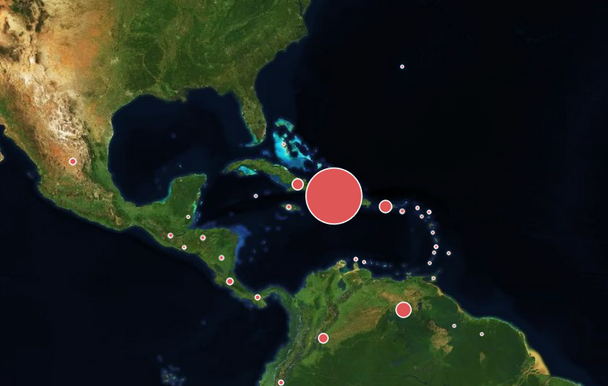
1. From the OSF repository (<https://osf.io/ekdr3/>) download: “*KNOMAD\_Bilateral\_Migration\_Matrix\_2017\_2018.xlsx”*
2. Save this file to a location you will remember--or create a new folder for it.
3. Open Tableau Public.
4. After you’ve opened Tableau Public, navigate to the top left “File” and select “Open.” Navigate to your “*KNOMAD\_Inward\_Remittance\_Flows\_2020.xlsx*” file, select it, and press Open.
5. For this data set, we’ll use the *“Data Interpreter”* function of Tableau, seen here:  
     
   
6. **NOTE: *“Data Interpreter”* is a useful tool, but it’s no substitute for cleaning the data yourself--what’s wrong with how this data set was cleaned?  
   **
7. Relabel “Destination country (across)” to “Origin country”
8. Go to your Worksheet in the bottom left and experiment with different country data and visualizations for Exercise #2.1. Here’s an example of what your “Marks” box might look like for one such visualization, which would visualize the Origin Country of migrants to the Dominican Republic, and an idea for what an early visualization of that data could look like:



  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Below are some incomplete example visualizations for Exercise #2.1--which scenario in Exercise #2.1 do you think these visualizations might apply to? Do you notice any glaring issues with these visualizations? Are they telling a compelling story?





\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**2.2 Exercise #2.2**

Collaborative Section:

These questions will be similar to the scenarios above, but this time, you’ll have multiple members in your group collaborating on the visualization. Each member will have their own separate agenda--the goal of the exercise will be to create the most ‘acceptable’ visualization that meets each individual members’ policy goals.

This exercise works best with groups of 2-4 members. Each group will select a “role” each with separate policy goals. Your group will work together to create a visualization with the provided (or external) data that best aligns with your overall policy goals--even if they don’t overlap nicely.

Example scenario: Working with remittance inflow data, your Minister of Development (Member A) wants to push the narrative that their country isn’t receiving enough foreign aid. However, they’re at the “top of the list” for remittance inflows, and other countries consider remittances as a major source of development. Your Defense Minister (Member B) wants to downplay how many remittances they’re receiving--they’re worried that high *formal* remittance inflows will be associated with high *informal* inflows, which have been linked to funding domestic terrorism. How can you reconcile these goals into one visualization? What common grounds can you find in the scenario, and can you create a unified “story” with them? Present your collaborative visualization to the class, explaining your reasoning behind the visual and narrative decisions you made as a group. (Use the Data Workflow from earlier in the Module to help.)

*Note: You may use any country-level data for this scenario, but if you plan to do the “competition” in Exercise 2.3, make sure that each group is using the same country data.   
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

**2.2 Exercise #2.3 (optional, see note above)**

“Competition”

For a fun exercise, choose one student (or your teacher) to act as the President of your scenario’s chosen country. The President has the final say in deciding which visualization best meets their political aims. The President should explain why they selected their ultimate “winner”--it may come down to their aesthetic preferences, visual clarity, or narrative power. It may simply be their whim. Good luck!

***\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_***

***2.3 Useful resources or “For Further Analysis or Discussion”***

For further data or resources on migration, check out the following sites: KNOMAD (<https://www.knomad.org/>), IOM (<https://www.iom.int/>), the Missing Migrants Project, (<https://missingmigrants.iom.int/>), and the UNHCR Data hub (<https://www.unhcr.org/en-us/data.html>).

For Tableau training videos, visit: <https://www.tableau.com/learn/training/20201>, or type “Tableau Tutorials” into your preferred search engine--there are also some excellent tutorials available on YouTube..

If available, you can also try the link below for Tableau’s own written tutorial: “Tutorial: Get Started with Tableau Desktop”

<https://help.tableau.com/current/guides/get-started-tutorial/en-us/get-started-tutorial-connect.htm>

Or the link below for their video training series: “Tableau Free Training Videos”

<https://www.tableau.com/learn/training/20201> ***\_\_\_\_\_\_\_\_***

Links to KNOMAD data on OSF/GitHub: <https://osf.io/ekdr3/>

(If any issues accessing, the same data is available at: <https://www.knomad.org/data/migration/emigration>

<https://www.knomad.org/data/remittances> )

***Bibliography:***

Packt. “Data Visualization: A Successful Design Process.”.<https://www.packtpub.com/product/data-visualization-a-successful-design-process/9781849693462>.

Wallace, Jeremy L. “Juking the Stats? Authoritarian Information Problems in China.” *British Journal of Political Science* 46, no. 1 (January 2016): 11–29.<https://doi.org/10.1017/S0007123414000106>.

Wickham, Hadley. “Tidy Data.” Journal of Statistical Software 59, no. 1 (September 12, 2014): 1–23.<https://doi.org/10.18637/jss.v059.i10>.

1. “Data Visualization: A Successful Design Process,” Packt, <https://www.packtpub.com/product/data-visualization-a-successful-design-process/9781849693462>. [↑](#footnote-ref-0)
2. Ibid. [↑](#footnote-ref-1)
3. Hadley Wickham, “Tidy Data,” *Journal of Statistical Software* 59, no. 1 (September 12, 2014): 1–23,<https://doi.org/10.18637/jss.v059.i10>. [↑](#footnote-ref-2)
4. Ibid. [↑](#footnote-ref-3)
5. Ibid. [↑](#footnote-ref-4)
6. Jeremy L. Wallace, “Juking the Stats? Authoritarian Information Problems in China,” *British Journal of Political Science* 46, no. 1 (January 2016): 11–29,<https://doi.org/10.1017/S0007123414000106>. [↑](#footnote-ref-5)