This document provides information on replication data for: A. François, S. Panel, L. Weill (2022), "Dictators' facial characteristics and foreign direct investment", *Leadership Quarterly* (forthcoming), https://doi.org/10.1016/j.leaqua.2022.101644. The replication package includes:

- A raw data file (FacialJudgementsData.dta) containing data from a survey admnistered between December 2021 and January 2022. The dataset includes facial judgements for 276 nondemocratic leaders in office from 1975 to 2010.
- The full dataset (including control variables) used for the analyses presented in the paper (ReplicationData.dta)
- A do file for replicating the results reported in the paper (ReplicationCode.do).

All files are in Stata version 14 (.dta) format.

When using the data, please cite: A. François, S. Panel, L. Weill (2022), "Dictators' facial characteristics and foreign direct investment", *Leadership Quarterly* (forthcoming)

## GENERAL INFORMATION

The raw data file (FacialJudgementsData.dta) contains facial judgements (trustworthiness, competence, attractiveness, maturity and masculinity) for 276 nondemocratic leaders in office from 1975 to 2010. Facial impressions were collected via an online survey administered in December 2021 and January 2022. There is one picture per leader, and one observation per leader/picture. Each of the 276 observations has a unique identifier ("pictureid") which is the same as in the full dataset. For each observation, we also provide the leader's name and a leader identifier drawn from Geddes et al. (2018), in addition to the country name and country COW identifier.

The full dataset (ReplicationData.dta) includes 2,895 observations (113 countries, 276 leaders) for the period 1975-2010. It includes only nondemocratic regimes, and only leaders who spent at least one full year in power. The observation is the country-leader-year, meaning that country-years in which a change of leadership took place appear twice in the dataset. The panel is unbalanced due to newly independent countries (e.g., Azerbaijan), countries that ceased to exist (e.g., East Germany), or countries switching back and forth between democracy and autocracy during the investigation period.

The replication dataset is organized in five parts. The three first parts (country and leader identifiers, regime and leader data, macroeconomic variables and other country-level covariates) include secondary data from publicly available sources. These variables appear only in the replication dataset. The two last parts (facial judgements and picture identifiers and characteristics) include data collected or coded by the authors. These variables are included both in the replication dataset and in the raw data file.

All variables are listed below with their definition and, when applicable, their source.

# VARIABLE LIST

## COUNTRY AND LEADER IDENTIFIERS

Secondary data (replication dataset only).

- ccode: Correlates of War country identifier
- Country: country name (string)
- year: 1975 to 2010
- leader: leader name as reported in Goemans et al. (2009)
- leaderid: numeric leader identifier (from Geddes et al. 2018).
- leadid: string leader identifier (from Goemans et al. 2009)
- transitionyear: takes on the value of 1 or 2 during years of leadership changes, 0 otherwise. 1 flags the leader in power on January 1st (the outgoing leader), 2 flags the leader in power on December 31 (the incoming leader). Note that the dataset only includes leaders who have spent at least one year in power, therefore there cannot be more than two leaders during one year.

Sources:
Geddes, B., Wright, J., and Frantz, E. (2018). How dictatorships work: Power, personalization, and collapse. Cambridge University Press. Data retrieved from: https://sites.psu.edu/dictators/how-dictatorships-work/
Goemans, H., Gleditsch, K.S., and Chiozza, G. (2009). Introducing *Archigos*: A Data Set of Political Leaders, *Journal of Peace Research*, 46(2): 269-183. Data retrieved from: https://www.rochester.edu/college/faculty/hgoemans/data.htm

## REGIME AND LEADER DATA

Secondary data (replication dataset only).

- eindate: day in which the leader entered office (from Goemans et al. 2009)
- eoutdate: day in which the leader left office (from Goemans et al. 2009)
- einyear: year in which the leader took office
- eoutyear: year in which the leader left office
- yrborn: leader's year of birth (from Goemans et al. 2009)
- ageleader: age of the leader in the current year
- tenurelength: records the total number of years the leader has been in power during the investigation period.
- duration: running sum of the number of years the leader has been in power
- entry: string variable indicating the manner in which the leader took office: regular, irregular, or foreign imposition (from Goemans et al. 2009).
- hereditary: 1 if the leader inherited power from relatives in a monarchy (from Geddes et al. 2018; original variable name is ldr_group_hereditary)
- electldr_multielec: dummy variable identifying leaders who were directly elected in an authoritarian regime allowing multiparty competition (from Geddes et al. 2018).
- electldr_multileg: dummy variable identifying leaders who were indirectly elected in an

authoritarian regime allowing multiparty competition (from Geddes et al. 2018).

- electldr_priordem: dummy variable identifying leaders who were elected in a previously democratic regime (from Geddes et al. 2018)
- elected: 1 if the leader was elected in a previously democratic regime or in an authoritarian regime allowing multiparty competition. Built from Geddes et al. (2018), variables electldr_multileg, electldr_multielec, electldr_priordem. The variable only takes into account the way in which the leader first entered office and codes as 0 those who first seized power via a coup, a rebellion, or a foreign imposition, and organized elections later during their term (e.g., Stroessner, Paul Kagame...). Thus, the value of the variable is constant at the leader level.
- ldr_group_insurgency: 1 if the leader took power as a result of a successful armed rebellion (from Geddes et al. 2018).
- irregular: 1 if the leader took office in a nonconstitutional manner (coup, rebellion, or foreign imposition). Built from the variable "entry" in Goemans et al. (2009) to identify coups and foreign impositions and the variable "ldr_group_insurgency" in Geddes et al. (2018) to identify insurgencies.

Sources: Geddes et al. (2018), Goemans et al. (2009).

MACROECONOMIC VARIABLES AND OTHER COUNTRY-LEVEL COVARIATES

Secondary data (replication dataset only).

- inflation: inflation as measured by the GDP deflator, in % (World Bank 2022)
- deflator: GDP deflator (built from variable "inflation")
- FDI: foreign direct investment net inflows in current USD (World Bank 2022)
- FDI_constant: FDI/deflator
- population: total population (World Bank 2022).
- FDI_population: annual FDI inflows per capita in constant USD (FDI_constant/population)
- log_fdi: log transformation of FDI_constant. A constant value of 1 USD was added to all observations in order to deal with null values. The absolute value of negative amounts was logged before restoring the negative sign.
- gdp: gross domestic product in 2015 USD (World Bank 2022)
- loggdp: log transformation of gdp
- trade: sum of imports and exports as a share of GDP (World Bank 2022)
- rents: share of natural resource rents in GDP (World Bank 2022)
- commonlaw: dummy variable taking the value of 1 for countries with a common law legal tradition, 0 otherwise (La Porta et al. 2008)
- war: dummy variable taking a value of 1 for each year in which the country was involved in an interstate conflict resulting in at least 1,000 battle-related deaths (Maoz et al. 2019).
- campaign: dummy variable taking a value of 1 for each year in which the country experienced a violent or nonviolent civil resistance campaign involving at least 1,000 participants (Chenoweth and Shay 2020).
- expend: military expenditures per soldier. The ratio was built from the variables "milex" (military expenditures) and "milper" (military personnel) in the National Material Capabilities dataset of the Correlates of War project.

Sources:
Chenoweth, E., and Shay, C.W. (2020), List of campaigns in NAVCO 1.3. Harvard dataverse. Data retrieved from https://dataverse.harvard.edu/dataverse/navco
Correlates of War (2021), National Material Capabilities (v6.0) dataset. Data retrieved from https://correlatesofwar.org/data-sets/national-material-capabilities
La Porta, R., Lopez-de Silanes, F., and Shleifer, A. (2008). The economic consequences of legal origins. *Journal of Economic Literature*, 46(2): 285–332. Data retrieved from https://faculty.tuck.dartmouth.edu/rafael-laporta/research-publications
Maoz, Z., Johnson, P. L., Kaplan, J., Ogunkoya, F., and Shreve, A. P. (2019). The dyadic militarized interstate disputes (MIDs) dataset version 3.0: Logic, characteristics, and comparisons to alternative datasets. *Journal of Conflict Resolution*, 63(3): 811–835. Data retrieved from https://correlatesofwar.org/data-sets/MIDs/dyadic_mid_4-01.zip/view
World Bank (2022), World Development Indicators. Data retrieved from https://data.worldbank.org/indicator/

FACIAL JUDGEMENTS

Source: authors' data. Variables appear in the two datasets.

We collected facial judgements for 276 pictures of nondemocratic leaders in office from 1975 to 2010. Data was collected for five dimensions (trustworthiness, competence, attractiveness, maturity, and masculinity, rated on a scale from 1 to 8) and was subject to three quality checks: 1) we implemented an Instructional manipulation check and removed answers from respondents who did not pass the test; 2) we screened deviant answers; 3) we removed answers from respondents who claimed they could identify the individual on a photograph (see details in François, Panel & Weill, 2022).

Variables are named as follows:

- All variables include a stem ("trust," "compet," "attrac," "matur" and "mascul") that identifies the dimension of interest: trusworthiness, competence, attractiveness, maturity and masculinity.
- Variables starting with the prefix "ave_" report the average score for each picture and dimension (for example, variables starting with "ave_trust" report the average trusworthiness score for a picture). Variables starting with the prefix "sd_" report the standard deviation, and variables starting with the prefix "eff_" report the number of valid responses.
- The suffix "_t1" means that the score was computed based on answers from respondents who passed the instructional manipulation check. The suffix "_t2" means that the score was computed after eliminating deviant answers. The suffix "_t3" means that answers from respondents who claimed they recognized the subject were removed. Thus, variables ending with the suffix "_t13" report summary statistics for respondents who passed the catch trial question AND did not recognize the subject; variables ending with the suffix "_t123" report summary statistics for respondents who passed the catch trial, did not give a deviant answer and did not recognize the subject; etc.
- Variables including the prefix "diff_" indicate that the score on the dimension of interest is not expressed in absolute value but relative to a benchmark face which was included in all questionnaires and subject to the same evaluations as leaders' faces. To compute the adjusted

score, we substracted the benchmark's score from each dictator's score (negative values thus mean that the dictator received a lower score than the benchmark on the dimension of interest) and then computed the average of valid responses.

Examples:

- The variable ave_attrac_t13 gives the average attractiveness score for each picture; the average includes only respondents who passed the instructional manipulation check (condition 1) and did not recognize the subject (condition 3).
- The variable eff_matur_t1 reports, for each picture, the effective number of valid responses on the maturity dimension after removing respondents who failed the catch trial.
- The main predictors used in the paper, ave_diff_trust_t13 and ave_diff_compet_t13, respectively report the differences in trustworthiness and competence ratings between each dictator's picture and the benchmark face, averaged over over all respondents who successfully passed the catch trial question and did not recognize the subject. The variables ave_diff_trust_t123 and ave_diff_compet_t123 measure the same thing except that we additionnally exclude respondents who gave a deviant answer from the average.

PICTURE IDENTIFIERS AND CHARACTERISTICS

Source: authors' data. Variables appear in the two datasets.

- pictureid: numeric picture identifier
- ave_qual_t13 is a measure (on a scale from 1 to 8) of the quality of the image, rated by survey participants. The suffix "_t13" means we only kept scores given by respondents who passed the catch trial and did not recognize the subject on the picture.
- smiling: dummy variable taking the value of 1 if the leader smiles on the picture.
- gazedirection: dummy variable taking the value of 1 if the leader is shown looking at the camera, 0 otherwise.
- ethnicity: 0 if the leader is white, 1 if he is Asian, 2 if he is black.
- glasses: 1 if the leader wears glasses on the picture, 0 otherwise.
- facialhair: dummy variable taking the value of 1 if the leader has either a beard or a moustache on the photograph, 0 otherwise.
- bald: dummy variable taking the value of 1 if the leader is completely bald on the picture, 0 otherwise.
- picturedate: year in which the picture was taken (only reported for long-serving rulers, tenurelength $\geq$ 15).
- window: dummy variable taking a value of 1 if more than 15 years have passed since the picture was taken. The variable takes the value of 0 if the leader spent less than 15 years in office or the leader spent 15 years or more in office but the picture was taken less than 15 years ago.