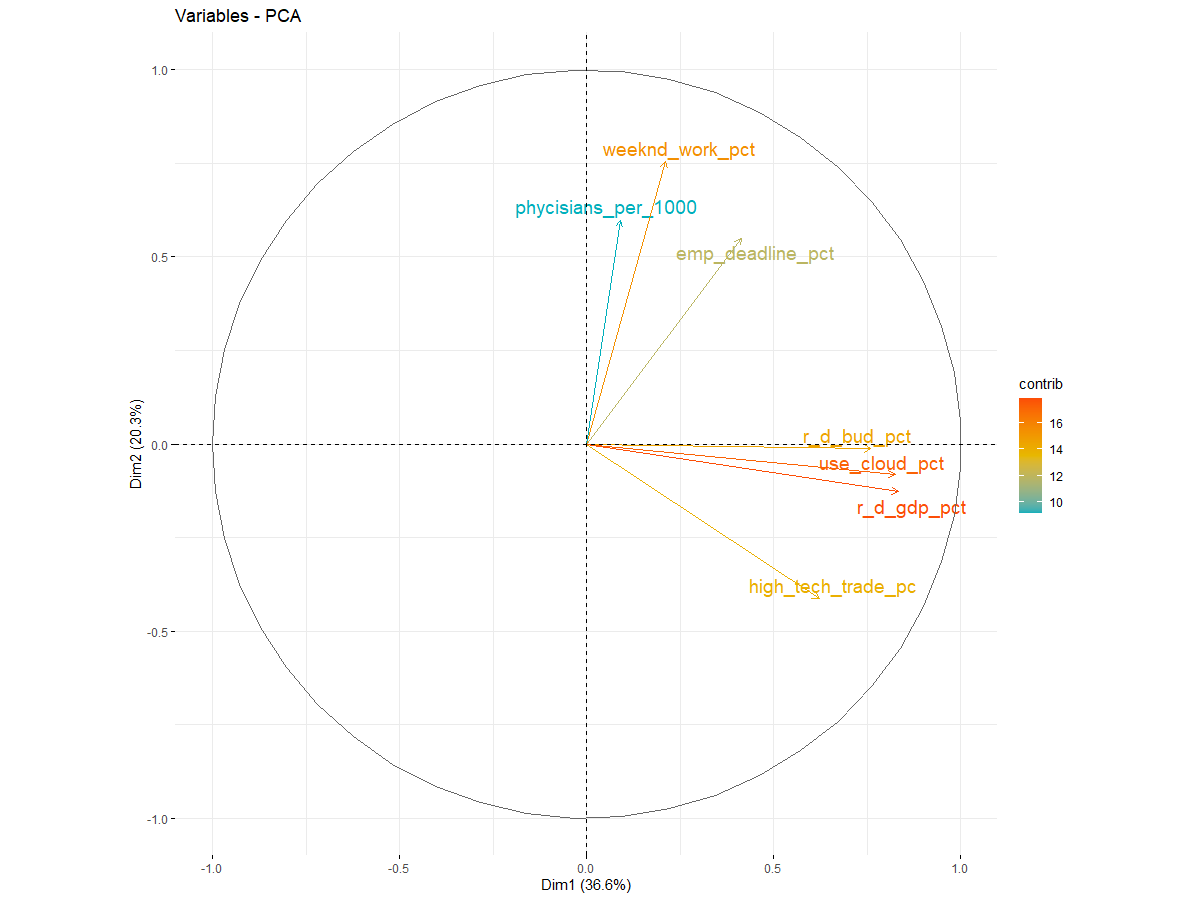
Efekt 7, Regresja z użyciem wymiarów

Mateusz Nowak, Damian Okoń, Robert Zamiar

# Pytanie badawcze

Czy korzystając z nowych zmiennych powstałych przy pomocy przeprowadzonej analizy PCA, uproszczony model regresji liniowej ma dużo gorsze dopasowanie? Czy ten model nadaje się do predykcji?

# Wstęp do analizy



Rysunek 1. Analiza PCA - wykres ładunków czynnikowych

W wyniku redukcji wymiarów otrzymano dwie składowe – PC2 opisującą kraje pod kątem społecznym, oraz PC1, w skład której weszły predyktory odpowiedzialne za nakłady pieniężne na technologię oraz szeroko rozumiany postęp technologiczny. Dwie zmienne z pierwszego wymiaru (high\_tech\_trade\_pc i r.d\_gdp\_pct) oraz jedna zmienna z drugiego wymiaru – weekend\_work\_pct były istotnymi predyktorami w stworzonych przez nas modelach regresji liniowej.

Obraz zawierający tekst, paragon

Opis wygenerowany automatycznie

Rysunek 2. Wyniki modelu regresji liniowej

# Analiza modelu regresji liniowej z jednym wymiarem

Do wstępnej analizy użyto wymiaru, w skład którego wchodzą zmienne związane z sprawami rozwojowymi państw.

Obraz zawierający tekst, paragon, zrzut ekranu

Opis wygenerowany automatycznie

Rysunek 3. Model regresji liniowej z jednym wymiarem

Pierwszy wymiar wyjaśnia zmienną objaśnianą (PKB per capita) w 79%. Zarówno predyktor, jak i wyraz wolny okazały się istotnie wpływać na zmienną celu. Uzyskany model jest istotny – p-value  
wynosi 0.

# Analiza modelu regresji liniowej z dwoma wymiarami

Obraz zawierający tekst, paragon, zrzut ekranu

Opis wygenerowany automatycznie

Rysunek 4. Model regresji liniowej z dwoma wymiarami

W modelu regresji liniowej z dwoma wymiarami predyktorem nieistotnie wpływającym na zmienną objaśnianą jest wymiar opisujący kraje pod względem polityki społecznej. W porównaniu do modelu z poprzedniego punktu należy zauważyć, że różnica pomiędzy R2, a skorygowanym R2 powiększyła się. Do modelu dodano zatem nieistotny predyktor, jakim jest PC2. Jeżeli chodzi o te miary, warto zauważyć, że współczynnik determinacji z poprzedniego punktu jest praktycznie taki sam jak w modelu z dwoma wymiarami.

# Statystyczne porównanie modeli

Obraz zawierający tekst

Opis wygenerowany automatycznie

Rysunek 5. Test ANOVA

Test wykonano przy hipotezie zerowej mówiącej, że oba modele są równe przydatne w celach predykcyjnych i H1 mówiącej o tym, że lepszy wydaje się większy model – z dwoma wymiarami. Otrzymana statystyka testowa jest większa niż poziom istotności, co oznacza, że nie ma podstaw do odrzucenia hipotezy zerowej. Do predykcji nadają się dwa wykonane przez nas modele. Lepszym rozwiązaniem jednak, zgodnym z prawem ekonometrii byłoby wykorzystanie pierwszego modelu.

# Odpowiedź na pytanie badawcze

Uzyskany model regresji liniowej z użyciem pierwszego wymiaru przedstawia się następująco:  
**y = 23677 + 11664 \* PC1** i charakteryzuje się współczynnikiem determinacji na poziomie 79%, identycznym do modelu regresji wykonanego wcześniej na samych predyktorach. Dzięki przekształceniu zmiennych przy pomocy PCA i rozłożeniu ich na wymiary otrzymujemy proste równanie, w którym zamiast 5 predyktorów mamy jeden. Taki uproszczony model może równie dobrze posłużyć do przewidywania wartości PKB per capita, korzystając z danych związanych z rozwojem ekonomiczno-technologicznym danego państwa.