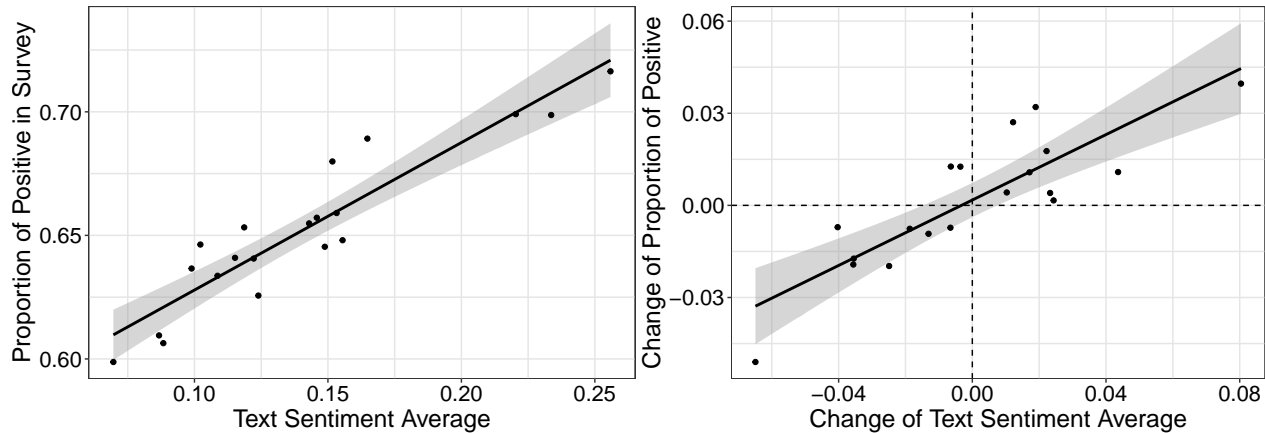# Materials for 'Validating daily social media macroscopes of emotions'

Der Standard

Scatterplot regression and correlation

```
##
## Call:
## lm(formula = prop_good ~ sent_roll3_weighted, data = pl)
##
## Residuals:
##       Min         1Q     Median         3Q        Max
## -0.0165216 -0.0105770 -0.0005588  0.0053255  0.0226208
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         0.568199   0.008432   67.39  < 2e-16 ***
## sent_roll3_weighted 0.596803   0.056837   10.50 4.19e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01221 on 18 degrees of freedom
## Multiple R-squared:  0.8597, Adjusted R-squared:  0.8519
## F-statistic: 110.3 on 1 and 18 DF,  p-value: 4.192e-09

##
##  Pearson's product-moment correlation
##
## data:  pl$prop_good and pl$sent_roll3_weighted
## t = 10.5, df = 18, p-value = 4.192e-09
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
```

```
##   0.8218585 0.9712129
## sample estimates:
##       cor
## 0.9271754
```

Scatterplot test of coefficients

```
##
## t test of coefficients:
##
##                     Estimate Std. Error t value  Pr(>|t|)
## (Intercept)           0.568      0.011   53.154 < 2.2e-16 ***
## sent_roll3_weighted   0.597      0.063    9.523 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##                      2.5 % 97.5 %
## (Intercept)          0.546  0.591
## sent_roll3_weighted 0.465  0.728
```

Differences regression and correlation

```
##
## Call:
## lm(formula = difgood ~ difsent, data = pl)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.018175 -0.006874 -0.002035  0.008362  0.020188
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.001737   0.002647   0.656     0.52
## difsent     0.532768   0.080577   6.612  4.4e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01154 on 17 degrees of freedom
## Multiple R-squared:   0.72,  Adjusted R-squared:  0.7035
## F-statistic: 43.72 on 1 and 17 DF,  p-value: 4.404e-06

##
##  Pearson's product-moment correlation
##
## data:  pl$difgood and pl$difsent
## t = 6.6119, df = 17, p-value = 4.404e-06
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.6416117 0.9403298
## sample estimates:
##       cor
## 0.8485356
```
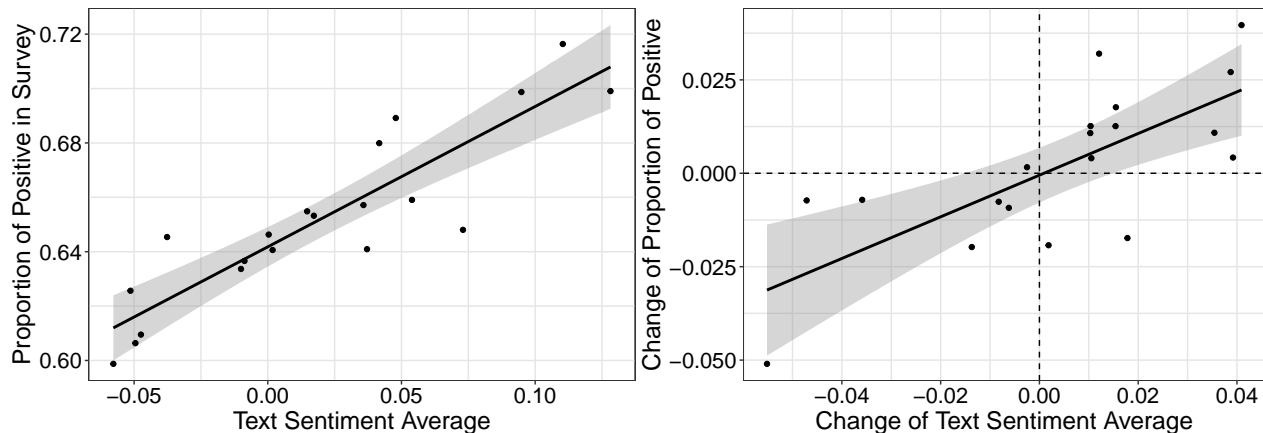
Differences test of coefficients

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.002      0.002    1.092     0.29
## difsent        0.533      0.068    7.885   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##               2.5 % 97.5 %
## (Intercept) -0.002  0.005
## difsent      0.390  0.675
```

Side by side plot
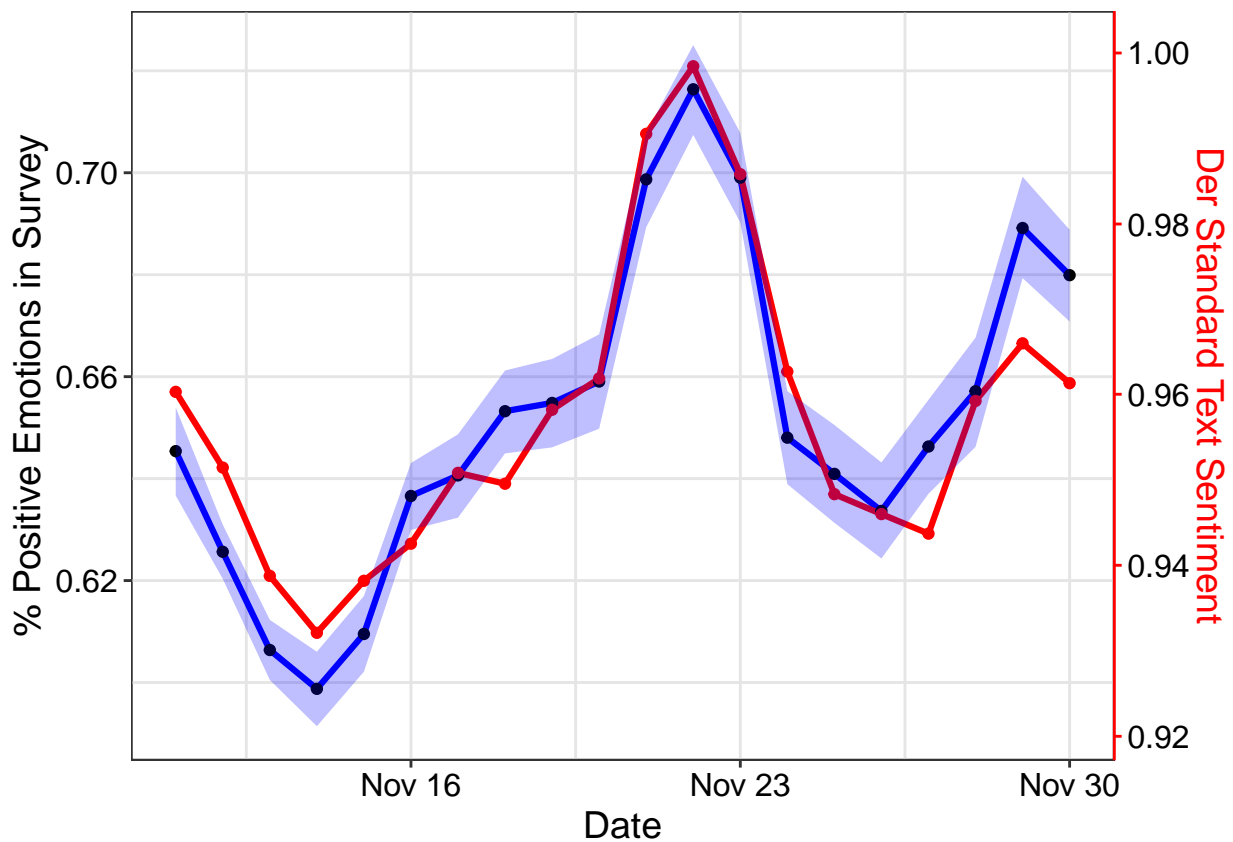
Twitter

Scatterplot regression and correlation

```
##
## Call:
## lm(formula = prop_good ~ sent_roll3_weighted_shifted1, data = cdplstandard)
##
## Residuals:
##       Min       1Q    Median       3Q       Max
## -0.031429 -0.009133 -0.001396  0.008562  0.023044
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  0.641786   0.003458 185.572  < 2e-16 ***
## sent_roll3_weighted_shifted1 0.516112   0.060520   8.528 9.75e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01452 on 18 degrees of freedom
## Multiple R-squared:  0.8016, Adjusted R-squared:  0.7906
## F-statistic: 72.73 on 1 and 18 DF,  p-value: 9.748e-08

##
##  Pearson's product-moment correlation
##
## data:  cdplstandard$prop_good and cdplstandard$sent_roll3_weighted_shifted1
## t = 8.528, df = 18, p-value = 9.748e-08
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.7499162 0.9582036
## sample estimates:
##       cor
## 0.8953214
```

Scatterplot test of coefficients

```
##
## t test of coefficients:
##
```

```
##                             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)                    0.642      0.004 167.248 < 2.2e-16 ***
## sent_roll3_weighted_shifted1   0.516      0.065   7.892 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##                              2.5 % 97.5 %
## (Intercept)                  0.634  0.650
## sent_roll3_weighted_shifted1 0.379  0.654
```

Differences regression and correlation

```
##
## Call:
## lm(formula = difgood ~ difsent, data = cdplstandard)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.026760 -0.010002  0.003522  0.008462  0.025795
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.000511   0.003473  -0.147 0.884748
## difsent      0.557438   0.127535   4.371 0.000416 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01496 on 17 degrees of freedom
## Multiple R-squared:  0.5291, Adjusted R-squared:  0.5014
## F-statistic:  19.1 on 1 and 17 DF,  p-value: 0.0004164

##
##  Pearson's product-moment correlation
##
## data:  cdplstandard$difgood and cdplstandard$difsent
## t = 4.3709, df = 17, p-value = 0.0004164
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.4080285 0.8881772
## sample estimates:
##       cor
## 0.7274241
```
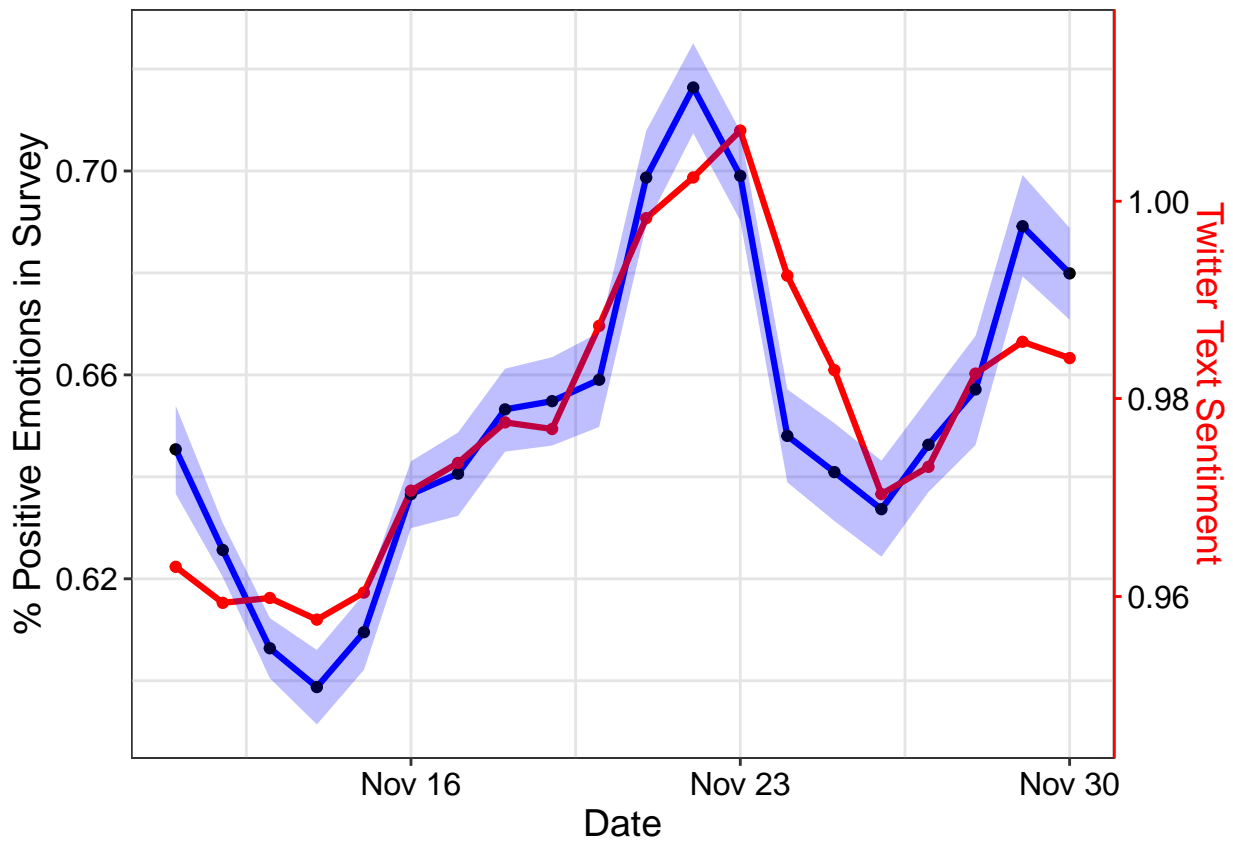
Differences test of coefficients

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -0.001      0.004  -0.133    0.896
## difsent         0.557      0.124   4.504   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##               2.5 % 97.5 %
## (Intercept) -0.009  0.008
```

```
## difsent      0.296  0.819
```

Side by side plot
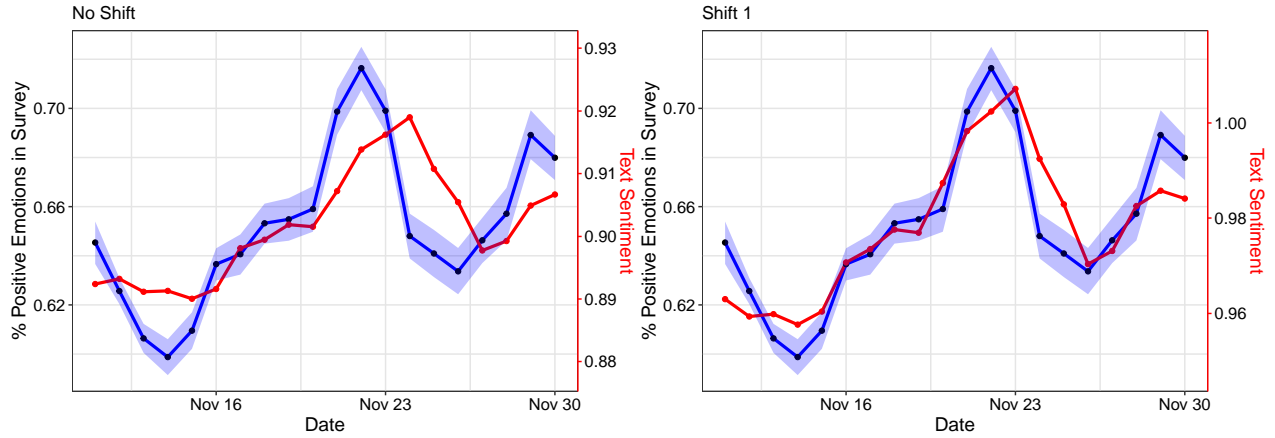

Twitter Pre-Registration Correlations

Exactly like in Pre-Registration

```
## [1] 0.63
```

```
## [1] "[0.26,0.84]"
```

```
## [1] 0.003
```

Change to follower limit of 100 000

```
## [1] 0.71
```

```
## [1] "[0.39,0.88]"
```

```
## [1] 0.00047
```

Twitter Shifts

Tables

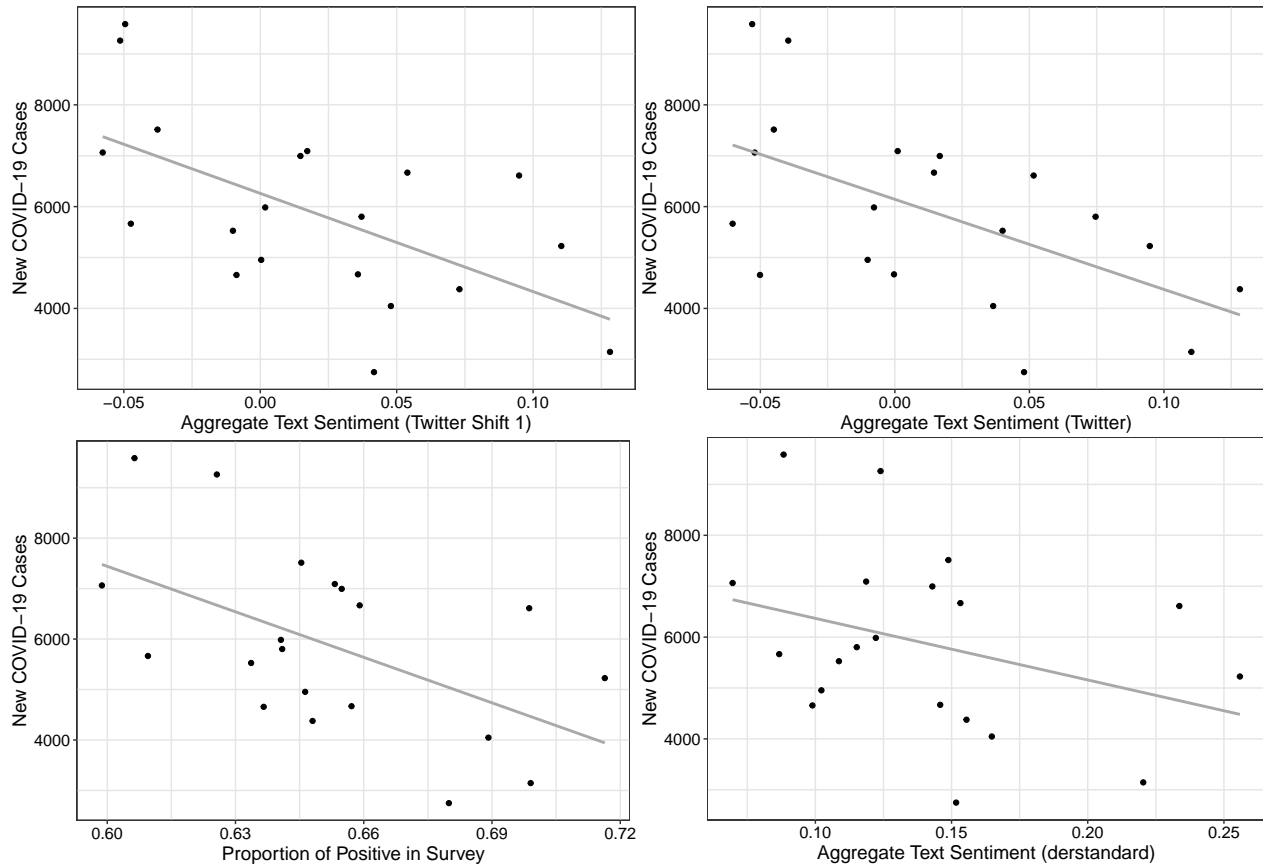|	derstandard	Twitter (Shift 1)	Twitter (No Shift)
LIWC+GS	0.93 [0.82,0.97]	0.90 [0.75,0.96]	0.71 [0.39,0.88]
LIWC	0.74 [0.44,0.89]	0.85 [0.64,0.94]	0.66 [0.31,0.85]
LIWC pos	0.81 [0.56,0.92]	0.80 [0.55,0.92]	0.80 [0.55,0.92]
LIWC neg	0.03 [-0.42,0.46]	-0.73 [-0.89,-0.43]	-0.73 [-0.89,-0.43]
GS	0.91 [0.78,0.96]	0.91 [0.79,0.97]	0.74 [0.43,0.89]
GS pos	0.89 [0.75,0.96]	0.91 [0.79,0.97]	0.91 [0.79,0.97]
GS neg	-0.57 [-0.81,-0.18]	-0.39 [-0.71,0.07]	-0.39 [-0.71,0.07]

	p Value	Absolute difference
Twitter vs Der Standard (Aggregate Sentiment)	0.4497163	0.03 (Der Standard)
GS vs. LIWC (Der Standard)	0.0495599	0.17 (GS)
GS vs. LIWC (Twitter)	0.1323296	0.07 (GS)
LIWC neg vs. LIWC pos (Der Standard)	0.0053294	0.78 (LIWC pos)
LIWC neg vs. LIWC pos (Twitter)	0.5370929	0.07 (LIWC pos)
GS neg vs. GS pos (Der Standard)	0.0094838	0.32 (GS pos)
GS neg vs. GS pos (Twitter)	0.0003683	0.53 (GS pos)

Evaluation with COVID-19 cases

	New Cases
Twitter (Aggregate Shift 1)	-0.60 [-0.82,-0.21]
Twitter (Aggregate No Shift)	-0.57 [-0.81,-0.17]
Survey	-0.53 [-0.79,-0.12]
Der Standard (Aggregate)	-0.33 [-0.68,0.13]

	p Value	Difference (higher)
Twitter (Aggregate Sentiment) vs. Survey	0.501	0.06 (Twitter)
Twitter (LIWC) vs. Survey	0.617	0.05 (Twitter LIWC)
Twitter (GS) vs. Survey	0.644	0.04 (Twitter GS)
derstandard (Aggregate Sentiment) vs. Survey	0.015	-0.2 (Survey)
derstandard (LIWC) vs. Survey	0.023	-0.36 (Survey)
derstandard (GS) vs. Survey	0.087	-0.16 (Survey)

Scatterplots COVID-19 cases

Bootstraping